May 1995

SU-ITP-94-16
RU-95-13
hep-ph/9405428

The Unified Minimal Supersymmetric Model with Large Yukawa Couplings

Riccardo Rattazzi[1]
Department of Physics and Astronomy
Rutgers University
Piscataway, NJ 08855

Uri Sarid[2]
Physics Department
Stanford University
Stanford, California 94305

Abstract

The consequences of assuming the third-generation Yukawa couplings are all large and comparable are studied in the context of the minimal supersymmetric extension of the standard model. General aspects of the RG evolution of the parameters, theoretical constraints needed to ensure proper electroweak symmetry breaking, and experimental and cosmological bounds on low-energy parameters are presented. We also present complete and exact semi-analytic solutions to the 1-loop RG equations. Focusing on SU(5) or SO(10) unification, we analyze the relationship between the top and bottom masses and the superspectrum, and the phenomenological implications of the GUT conditions on scalar masses. Future experimental measurements of the superspectrum and of the strong coupling will distinguish between various GUT-scale scenarios. And if present experimental knowledge is to be accounted for most naturally, a particular set of predictions is singled out.

[1]E-mail: rattazzi@physics.rutgers.edu
[2]E-mail: sarid@squirrel.stanford.edu. Address after July 1995: Dept. of Physics, University of Notre Dame, Notre Dame IN 46556.

1 Introduction

Unified theories are currently the most promising candidates for physics beyond the standard model. The marriage of force unification, namely grand unified theories (GUTs) or perhaps string theory, and spin unification, by which we mean a supersymmetry relating fermions and bosons, has been a fruitful and prolific area of research in the last few decades. Grand unification allows the understanding of the electroweak and strong forces as low-energy manifestations of a single microscopic force, in particular explaining the quantization and the assignments of electromagnetic charges for all known particles. The simplest GUTs [1, 2], based on SU(5) or SO(10) gauge groups, unify some [SU(5)] or all [SO(10)] of the quarks and leptons in each generation. The unified matter multiplets neatly encompass the known standard-model matter particles — no new particles are needed, and no known particles are left out. The one exception is the right-handed neutrino, which must be included in SO(10) unification. If neutrinos have masses, this potential embarrassment becomes a boon. In fact, SO(10) not only favors typical ranges for their masses via the seesaw mechanism, but in specific models can also lead to detailed predictions about the flavor structure of the mass matrix. Moreover, SO(10) beautifully incorporates both the Pati-Salam idea [3] of leptons as the fourth color and an underlying symmetry between the left- and right-handed quarks and leptons [4]. String theory aims to go beyond GUTs and unify all forces including gravity, perhaps without any adjustable parameters. Though still in their infancy, string models would presumably reproduce the successes of GUTs by either implying a grand unified theory as a "low-energy" consequence or by furnishing the appropriate boundary conditions to mimic grand unification predictions. GUTs, and even more so string theory, ambitiously span energy scales some 13 to 15 orders of magnitude above the highest scales at which the standard model has been directly tested. For compelling esthetic reasons such a span requires these theories to be supersymmetric: fermions and bosons, present in equal numbers, with mirror (and therefore highly restricted) interactions. Supersymmetric theories are theoretically attractive on their own, but when wedded to unified theories they can yield quantitative phenomenological successes. Supersymmetry (SUSY) cannot be an exact symmetry of nature, but it must be approximately valid down to roughly the electroweak scale if such a low scale is to have a hope of being naturally embedded in a GUT- or string-scale theory [5]. The minimal candidate model for unification is then given at low energies by the supersymmetrized standard model, which we will call the MSSM[3], having a squark for each quark, a slepton for each lepton, a gaugino for each gauge boson, and Higgsinos for the requisite two Higgs doublets H_u and H_d which give masses to the up- and down-type quarks and leptons (respectively). For a given effective theory below the unification scale, such as the MSSM, the renormalization group (RG) evolution determines the relationship between the physics of the unified theory and the physics observed at the electroweak scale. If one chooses as the effective theory the particular particle content of the MSSM, and embeds this minimal model in an SU(5) GUT,

[3]Here we use the term MSSM to refer only to the particle content and interactions of the minimal supersymmetric extension of the (essentially) experimentally established standard model, without any assumptions about the boundary conditions on its parameters at the GUT or string scales.

one arrives at a remarkably successful predicted relationship between the three low-energy gauge couplings [6, 7, 8]. There are also various predictions for the quarks and leptons, some more robust than others. In particular, the bottom quark and tau lepton Yukawa couplings are in most cases predicted to be equal at the GUT scale [9]; a successful prediction for the bottom mass at low energies, especially given a heavy top quark, can then be easily obtained with precisely the two Higgs doublets needed for the MSSM [10]. The large Yukawa coupling of the heavy top quark can also trigger a correct breakdown of electroweak symmetry at low energies [11, 12]. Thus the various pieces of the unification puzzle interlock tightly to produce a framework which we find both theoretically and phenomenologically compelling.

Numerous authors have explored in detail the issues associated with gauge coupling unification, in GUTs and lately also in string theories. The present paper takes a different path and seeks the consequences of Yukawa coupling unification [13, 14]. We will only be concerned with the Yukawa couplings of the third generation, namely the top and bottom quarks and the tau lepton, for two reasons: first, because they are larger than the Yukawas of the lighter generations, so it is natural to expect that they arise directly and from renormalizable operators and so are robustly predicted by the unified model, whereas the smaller couplings presumably arise from more complicated mechanisms and are thus more model-dependent; and second, because the third generation Yukawa couplings are the only ones big enough to appreciably influence the rest of the MSSM via the RG evolution. The focus of our research parallels these two motivations: we have seen in previous work [15], as summarised below, that the top mass can be predicted from approximate or exact unification at the GUT scale; and we expand on our previous observations [16] that the large bottom and tau Yukawa couplings which result from such unification qualitatively change the expected features of the MSSM at low energies.

The assumption underlying this work is that, at the unification scale, either (I) $\lambda_\tau = \lambda_b = \lambda_t$ or at least (II) $\lambda_\tau = \lambda_b \sim \lambda_t$, where $\lambda_{t,b,\tau}$ are the Yukawa couplings to the appropriate Higgs doublet, and "$\sim$" means that the couplings differ by a factor of order one. When is this assumption valid? In the simplest SO(10) scenario, in which both light Higgs doublets arise from a single SO(10) multiplet, the tree-level Yukawa couplings are exactly equal at the GUT scale, as in assumption (I) [17, 15]. Threshold corrections will typically lift this equality somewhat, and thereby can facilitate proper electroweak symmetry breaking, as we show below. In more involved SO(10) models, the light Higgs doublets may come from mixtures of several SO(10) multiplets. Nevertheless, we expect assumption (II) to often hold. In the simplest GUT scenarios based upon SU(5) the bottom and tau couplings are equal, but are unrelated to the top coupling. Most of the work on the MSSM has usually assumed that the top coupling was much larger than the other two, resulting in the observed hierarchy between the top and bottom quarks. From a GUT-scale model-building perspective it seems to us at least as natural *a priori* to assume that all three Yukawas are comparable, as in case (II); then the observed lightness of the bottom bottom and tau must result from the small vacuum expectation value (VEV) of the Higgs doublet to which they couple. Finally, the jury is still out on the predictions string theory makes for the Yukawa couplings. If the effective field theory which describes physics below the string scale is a GUT, then one of the above scenarios may hold [18]. Otherwise, there are still reasons to believe that the Yukawa

couplings for the third generation are approximately equal, at least in some string-inspired models.

In addition to the above more theoretical motivations, there is also a phenomenological astrophysical advantage to large $\tan\beta$, at least in SO(10) models [19, 20]. Various astrophysical and cosmological data, such as the neutrino solar flux deficit and the density fluctuations at large scale observed by COBE, can be explained if the left-handed neutrinos acquire a mass via the seesaw mechanism from right-handed neutrino Majorana masses in the $10^{10}-10^{13}\,\mathrm{GeV}$ range. A third-generation right-handed neutrino with such an intermediate mass can significantly affect bottom-tau unification (through the RG evolution of Yukawa couplings). If SO(10)-type boundary conditions $\lambda_b^G = \lambda_\tau^G$ and $\lambda_t^G = \lambda_N^G$ are assumed, it is difficult to reproduce the experimental value of m_b/m_τ for small $\tan\beta$ ($< 10-20$). Interestingly, though, for larger values of $\tan\beta$, the intermediate-scale right handed neutrino does not significantly alter the successful prediction of m_b/m_τ.

Motivated by the phenomenological successes of the unified MSSM and by the wide-ranging contexts displaying approximate or exact Yukawa unification, we analyze in this work various implications of a Yukawa-unified MSSM. In Sec. 2 we review the prediction of the top mass as a sensitive function not only of the GUT-scale boundary conditions and the bottom and tau masses, but also (perhaps surprisingly) of the superpartner masses. Reversing the argument, we find bounds on the superspectrum as functions of the top mass. Sec. 3 treats the related predictions for the radiative bottom quark decay $b \to s\gamma$ (and comments on $\tau \to \mu\gamma$). We point out the importance of seemingly subdominant diagrams which are also enhanced by large $\tan\beta$. Satisfying the recent experimental bounds on this process places certain constraints on the superspectrum if a delicate fine-tuning is to be avoided. In Sec. 4 we outline the basic implications of $\lambda_b \sim \lambda_t$ for electroweak symmetry breaking. Not only must the symmetry be broken radiatively without losing the $\mathrm{SU(3)_c} \times \mathrm{U(1)_{em}}$ gauge symmetries, but also a large hierarchy must be generated in the Higgs VEVs to account for the top-bottom mass hierarchy. Sec. 5 deals with the various options for generating this hierarchy. Two symmetries (PQ and R) can make this hierarchy more natural—and lead to a favored superspectrum—but there is always a necessary fine-tuning of at least one part in ~ 50 ($\sim m_t/m_b$) somewhere in the Lagrangian [21, 16]. We return in Sec. 6 to the problem of properly breaking the electroweak symmetry in the presence of the PQ and R symmetries. The various conditions which must be satisfied at appropriate scales to guarantee the proper spontaneous symmetry breaking are discussed in some detail in subsection 6.1. In addition to the various individual mass-squared parameters, we examine the two flat directions in the scalar potential (and the scales at which they may destabilize the vacuum) and the trilinear scalar couplings from A and μ terms. These can be important, even for the third generation, when there is some hierarchy between m_Z and the SUSY scale. We then turn to the general behavior of the soft scalar masses as they evolve down from the GUT scale, focusing in subsection 6.2 on the homogeneous part of the RG equations which dominates when the PQ and R symmetries are approximately valid. The favored scenarios with and without these symmetries are briefly summarized in subsection 6.3, without any assumptions about the soft scalar masses at the GUT scale. These assumptions are introduced in Sec. 7. The GUT-scale constraints on scalar masses in minimal SO(10)

theories, and in SU(5) or nonminimal SO(10) models, are presented in subsection 7.1 using a common notation. The ramifications of these SO(10)- or SU(5)-type GUT relations are explored in detail in subsections 7.2 and 7.3, respectively. These include the ease of obtaining proper electroweak symmetry breaking for different values of the parameters, the possible and probable superspectra, and the implications of lifting the PQ or R symmetries. We also study the effects of a right-handed neutrino mass below the GUT scale, briefly examine the consequences of universal scalar masses, and reconsider SU(5)-type boundary conditions as a perturbation on the SO(10)-type conditions. We then turn in Sec. 7.3 to the astrophysical and cosmological constraints on the Yukawa-unified MSSM. We address both the electric neutrality of the lightest superpartner and its relic abundance. To estimate this abundance, we adapt previous analyses to the large $\tan\beta$ scenario, and in the case of a bino-like LSP classify the dominant operators contributing to LSP annihilation in order to clarify its suppression. We present our conclusions in Sec. 9. In particular, we summarize the phenomenological expectations from Yukawa unification, comment on the most natural and therefore favored scenarios, and outline some directions for future investigation. In Appendix A we present the exact and complete semi-analytic solutions to the 1-loop RG equations for the MSSM with large $\lambda_{t,b,\tau}$. They are semi-analytic in that they are given in terms of integrals over the dimensionless (gauge and Yukawa) couplings, which must be evaluated numerically or approximated analytically, as we show for several examples. Appendix B is devoted to a study of one of the potentially flat directions in the scalar potential of the MSSM, and to the scales at which it can impose a constraint on the scalar mass parameters. Finally, Appendix C justifies the approximation we have made in using the RG-improved tree-level scalar potential.

2 Top mass: Prediction and Constraint

At tree level, the observed masses of the third family fermions are related to their Yukawa couplings and to the VEVs of the up- and down-type Higgs doublets via:

$$\begin{aligned} m_t &= \lambda_t v_U \equiv \lambda_t v \sin\beta \\ m_b &= \lambda_b v_D \equiv \lambda_b v \cos\beta \\ m_\tau &= \lambda_\tau v_D \equiv \lambda_\tau v \cos\beta \end{aligned} \tag{1}$$

where $v = 174\text{GeV}$ and $\tan\beta \equiv v_U/v_D \equiv \langle H_U^0\rangle/\langle H_D^0\rangle$. The Yukawa couplings are in turn determined through the renormalization group evolution by the Yukawa couplings $\lambda^G_{t,b,\tau}$ at the GUT scale $M_{\rm GUT}$. And finally, in the grand unified theory these couplings are related to each other according to:

$$\lambda^G_\tau = \lambda^G_b = \lambda^G_t \equiv \lambda_G\,, \tag{2}$$

or

$$\lambda^G_\tau = \lambda^G_b \sim \lambda^G_t \equiv \lambda_G\,. \tag{3}$$

In the minimal SO(10) scenario $\lambda^G_t/\lambda^G_b = 1$, while Higgs mixing or an SU(5) model could suggest that λ^G_t/λ^G_b is of order 1. The RG evolution requires as additional inputs the scale

of unification M_G and the unified value g_G of the gauge couplings, both of which are already fixed by gauge unification (but see Ref. [15] for the treatment of α_s). Thus the four low-energy observables m_t, m_b, m_τ and $\tan\beta$ and the two GUT-scale parameters λ_G and λ_t^G/λ_b^G are related by three (RG) equations; fixing m_b and m_τ from experiment leaves three equations in the four remaining variables m_t, $\tan\beta$, λ_G and λ_t^G/λ_b^G, yielding a single prediction for the top mass as a function of the angle β. (In principle, of course, the Higgs doublets VEVs v_U and v_D and hence also β are predicted in terms of the GUT-scale parameters of the Higgs sector, but at this stage those parameters are completely unknown; we will return to them below.) If, as in minimal SO(10), the initial ratio λ_t^G/λ_b^G is fixed, then the top mass and $\tan\beta$ are predicted individually. We have previously presented a detailed study of this prediction [15] when λ_t^G and λ_b^G are split at most by threshold corrections; we will return to the more general case $\lambda_t^G \sim \lambda_b^G$ below. A crucial finding of that work (valid more generally for any $\lambda_t^G \sim \lambda_b^G$) was related to the fact that, since the mass of the bottom quark results from a large coupling to a small VEV rather than a small coupling to a large VEV, any chiral symmetries protecting this mass should act on the VEV and not on the Yukawa coupling as in the usual case. Such approximate symmetries, discussed in detail below, are not respected by the parameters of the *generic* MSSM, which therefore exhibits large corrections to m_b from 1-loop couplings to the other Higgs VEV ([15]; see also [22]). In particular, we found that the usual suppression factor of $1/16\pi^2$ in the leading 1-loop corrections is *a priori completely undone* by the enhancement $\sim \tan\beta$ from the larger VEV of the up-type Higgs doublet. The two dominant contributions are given by the diagrams shown in Fig. 1. Keeping only these corrections (and therefore dropping similar but smaller corrections to m_τ), the third generation mass relations now read

$$\begin{aligned} m_t &= \lambda_t \left[\lambda_G, \lambda_t^G/\lambda_b^G\right] v \sin\beta \\ m_b &= \lambda_b \left[\lambda_G, \lambda_t^G/\lambda_b^G\right] v \cos\beta \left(1 + \frac{\delta m_b}{m_b}\right) \\ m_\tau &= \lambda_\tau \left[\lambda_G, \lambda_t^G/\lambda_b^G\right] v \cos\beta \end{aligned} \tag{4}$$

where we have explicitly shown the dependence of the three Yukawa couplings at low energies on the two GUT-scale parameters. The exact form of the corrections to the bottom mass was given in our previous work. A useful approximation is given by $\delta m_b/m_b = (\tan\beta/50)\ \delta_b$, where

$$\delta_b \simeq \frac{50}{16\pi^2}\frac{\alpha_G}{\alpha_W}\frac{\mu m_{\widetilde{W}}}{m_{\rm eff}^2}\left[\frac{8}{3}\frac{\alpha_s}{\alpha_G}g_s^2\, f\left(\frac{m_{\tilde{g}}^2}{m_{\rm eff}^2}\right) - 2\lambda_t^2\, f\left(\frac{\mu^2}{m_{\rm eff}^2}\right)\right], \tag{5}$$

$f(x) = (1 - x + x\ln x)/(1-x)^2$, $m_{\tilde{g}}$ and μ are the gluino mass and the μ term evaluated at the electroweak scale, $\alpha_W = g_2^2/4\pi$ and $\alpha_s = g_s^2/4\pi$ are the SU(2) and $\mathrm{SU(3)_c}$ coupling strengths respectively and $m_{\rm eff}^2 \equiv \frac{1}{2}(m_{\tilde{b}}^2 + m_{\widetilde{Q}}^2)$ is the average of the squared masses of the SU(2)-singlet bottom squark and the SU(2)-doublet third generation squarks. [In the second (subdominant) term we used $A_t \simeq 2M_{1/2} = 2\frac{\alpha_G}{\alpha_W}m_{\widetilde{W}}$ and approximated $m_{\tilde{b}}^2 + m_{\widetilde{Q}}^2 \simeq m_{\tilde{t}}^2 + m_{\widetilde{Q}}^2$. Also, the expressions are considerably modified when one of the stop or sbottom eigenvalues becomes very small, but this will not be relevant for the cases we study.] We see that even if

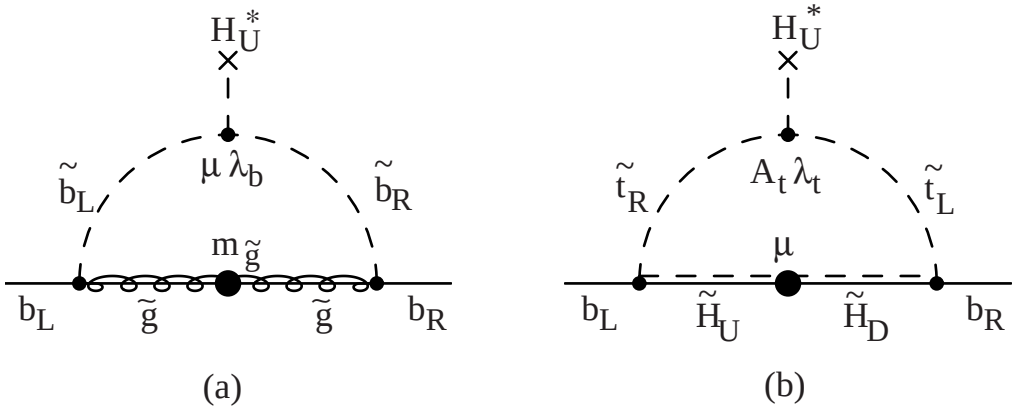

Fig. 1. The leading (finite) 1-loop MSSM contributions to the b quark mass.

λ_t^G/λ_b^G is fixed, for example in minimal SO(10), we now have the additional unknown δ_b, which precludes a separate determination of m_t and $\tan\beta$. In other words, if we don't know enough about the superspectrum to pin down δ_b, we cannot fix λ_G by comparing the prediction of m_b with experiment, and hence we lose the independent prediction of the top mass.

It will prove very helpful to understand the magnitude of the $\delta m_b/m_b$ corrections, and in fact of all large $\tan\beta$ enhancements, from the point of view of symmetries. To this end we recall the Peccei-Quinn (PQ) and R symmetries introduced in Ref. [15]. The PQ symmetry amounts to setting $\mu = 0$, while the R symmetry requires the vanishing of gaugino masses $M_{1/2}$, of the SUSY-breaking trilinear scalar couplings A_i, and of the bilinear SUSY-breaking Higgs coupling B. If either symmetry were exact, then when the up-type Higgs acquired a VEV the down-type VEV would remain zero, so $\tan\beta$ would be infinite. (Of course we have in mind the usual scenario in which the Higgs mass matrix has a negative eigenvalue in the H_u direction only.) Also, down-type quarks and leptons such as the bottom and tau would be exactly massless to all orders. We will see that these symmetries are the key to making large $\tan\beta$ as technically natural as possible: just as the bottom mass can be made as light as needed by imposing the usual chiral symmetry, so the PQ and R symmetries can be imposed on the Lagrangian to varying degrees. Unfortunately, current LEP bounds set strong bounds on how natural the large $\tan\beta$ scenario can be [16, 21], as we discuss below. But PQ and R are still the key to alleviating as much as possible the need for fine-tuning, and are also useful for classifying the various superspectra and discussing their phenomenological consequences.

If the symmetries are only approximately valid, one must specify at what scale this approximation holds: we will see that in very fine-tuned scenarios, the squark masses m_0 evaluated at the electroweak scale are much smaller than their values M_0 at the GUT scale, so an approximately symmetric GUT Lagrangian having $\mu \sim M_{1/2} \ll M_0$ could yield a spectrum at observable energies $\mu \sim M_{1/2} \sim m_0$ having no observable symmetries. Thus for those cases we will distinguish between having PQ and R symmetries at all scales, and having them only at high scales.

In Fig. 2, we present the results of a detailed 2-loop analysis, following Ref. [15], of the

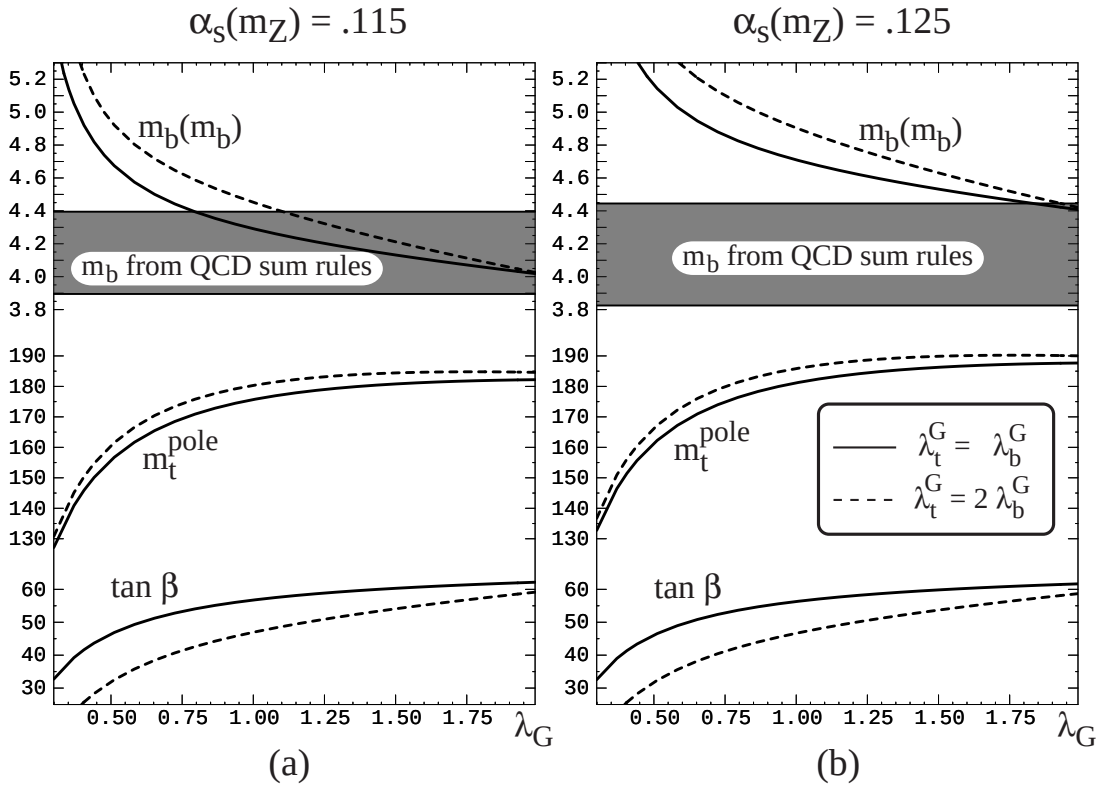

Fig. 2. The predictions of 2-loop RG evolution (plus 1-loop threshold corrections) for the running bottom quark mass (in GeV), the pole top quark mass (in GeV), and $\tan\beta$, as functions of the GUT-scale Yukawa coupling λ_G. To compute threshold corrections, we assumed the preferred superspectrum of Eq. (22). The solid lines correspond to exact Yukawa unification, while the dashed lines indicate $\lambda_G \equiv \lambda_t^G = 2\lambda_b^G$ as an example of approximate Yukawa unification. The values of m_b, after the finite corrections δm_b are added, should fall within the shaded bands in order to agree with the bottom mass as extracted from QCD sum rules [15].

running $\overline{\rm MS}$ bottom quark mass $m_b(m_b)$, the top quark pole mass, and the approximate ratio $\tan\beta$ as functions of λ_G within the perturbative regime. We chose two sample values of the QCD coupling $\alpha_s(m_Z)$, and considered either exact or approximate Yukawa unification, where the former is defined to be $\lambda_t^G = \lambda_b^G$ while the latter is exemplified by $\lambda_t^G/\lambda_b^G = 2$. (There is in addition a slight logarithmic dependence on the superspectrum, as calculated in Ref. [15]; for definiteness we have assumed the spectrum singled out below, namely all squarks and sleptons and the pseudoscalar Higgs at $\sim 600\,\text{GeV}$, while the SU(2) gaugino and the Higgsinos are at $\sim m_Z$.) Also shown are the corresponding allowed ranges for $m_b(m_b)$ extracted from the data on $e^+e^- \to b\bar{b}$ using QCD sum rules [23]. We use the values obtained in the analysis of Ref. [15], which update those in Ref. [24]. The uncertainty on $m_b(m_b)$ is essentially theoretical, being dominated by our ignorance of $\mathcal{O}(\alpha_s^2)$ corrections to QCD sum rules. In the absence of

$\lambda_t^G = \lambda_b^G$						
$m_t^{\rm pole}$	$\alpha_s(m_Z) = .115$			$\alpha_s(m_Z) = .125$		
GeV	λ_G	$\frac{\delta m_b}{m_b}$ (max,min)	δ_b (max,min)	λ_G	$\frac{\delta m_b}{m_b}$ (max,min)	δ_b (max,min)
150	.44	$(-.09, -.19)$	$(-.10, -.22)$	.39	$(-.19, -.30)$	$(-.23, -.37)$
160	.56	$(-.05, -.15)$	$(-.05, -.16)$	.48	$(-.14, -.26)$	$(-.16, -.29)$
170	.77	$(\ \ 0, -.12)$	$(\ \ 0, -.11)$	.63	$(-.11, -.23)$	$(-.10, -.23)$
180	1.4	$(\ .05, -.07)$	$(\ .05, -.05)$	.94	$(-.06, -.19)$	$(-.06, -.17)$
$\lambda_t^G = 2\lambda_b^G$						
$m_t^{\rm pole}$	$\alpha_s(m_Z) = .115$			$\alpha_s(m_Z) = .125$		
GeV	λ_G	$\frac{\delta m_b}{m_b}$ (max,min)	δ_b (max,min)	λ_G	$\frac{\delta m_b}{m_b}$ (max,min)	δ_b (max,min)
150	.40	$(-.16, -.26)$	$(-.32, -.50)$	.36	$(-.27, -.37)$	$(-.60, -.84)$
160	.49	$(-.11, -.21)$	$(-.18, -.34)$	.44	$(-.22, -.32)$	$(-.39, -.58)$
170	.64	$(-.07, -.18)$	$(-.09, -.23)$	.55	$(-.17, -.29)$	$(-.25, -.42)$
180	.98	$(-.02, -.13)$	$(-.02, -.14)$	.75	$(-.13, -.25)$	$(-.16, -.30)$

Table 1: The consequences of Yukawa unfication using 2-loop RG evolution and 1-loop logarithmic threshold corrections from the preferred spectrum of Eq. (22). For every value of the top quark pole mass we list the unified Yukawa coupling at the GUT scale, the minimim and maximum values of $\delta m_b/m_b$ needed to bring the bottom quark prediction into agreement with experimental data, and the corresponding values of the $\tan\beta$-independent quantity δ_b.

δ_b, we could read off an allowed range of λ_G by requiring agreement between the theoretical and experimental values of m_b; then both m_t and $\tan\beta$ could be predicted within some range. Instead, we can only determine the top mass (and $\tan\beta$) as functions of δ_b.

Turning the argument around, for a given top mass we can calculate the amount of correction $\frac{\delta m_b}{m_b}$ needed to bring the bottom mass prediction into agreement with experiment. We can then remove the $\tan\beta$ dependence, leaving only the spectrum-dependent quantity δ_b. Table 1 displays these minimal and maximal allowed values of $\frac{\delta m_b}{m_b}$ and δ_b. (Actually, these bounds on $\frac{\delta m_b}{m_b}$ and δ_b themselves depend on the spectrum due to threshold corrections, but this dependence is a weak logarithmic one; typically, the logarithmic variation in m_t is at most a few GeV for the more interesting higher values of m_t. To obtain precise predictions, though, all thresholds corrections should be included using a definite superspectrum.) We learn that a positive δ_b must be quite small, while a negative δ_b may be large enough in magnitude to bring high predictions of m_b back into agreement with experiment. For example, when $\alpha_s(m_Z) = .115$ and $\lambda_t^G = \lambda_b^G$, superpartner spectra for which $|\delta_b| \gtrsim 5\%$ *allow* a light top quark, whereas spectra for which $|\delta_b| \gtrsim 16\%$ *mandate* a light top, where by light we mean $m_t \lesssim 160\,\text{GeV}$. Conversely, when $\lambda_t^G = 2\lambda_b^G$, a superspectrum for which $|\delta_b| \lesssim 15\%$ mandates a heavy top and favors a small $\alpha_s(m_Z)$. Fig. 3 translates this information into constraints on $m_{\widetilde{W}}$ (the mass of the gaugino superpartner of the W) and μ at the electroweak scale,

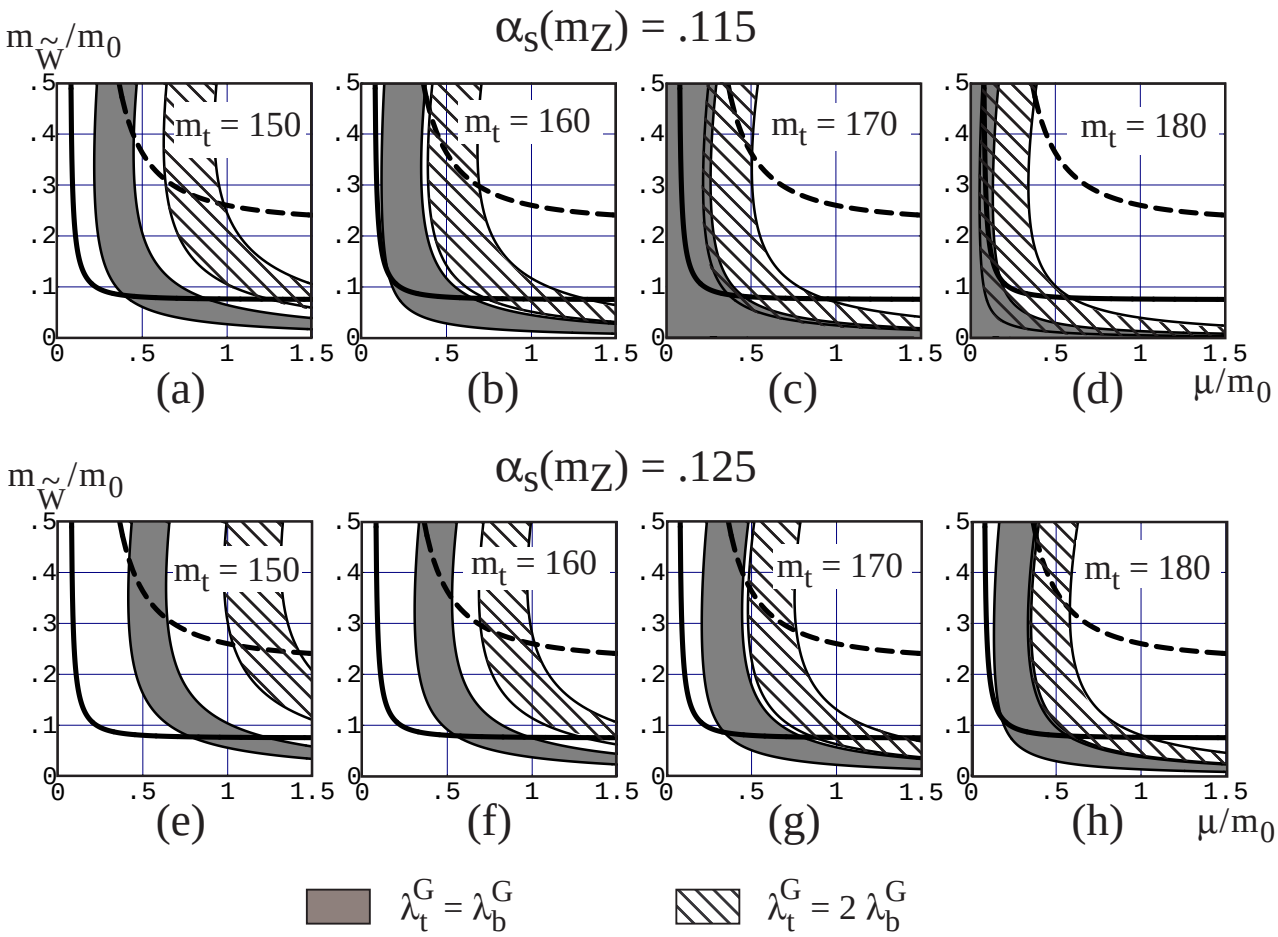

Fig. 3. The values of μ and the wino mass, normalized to a typical squark mass (specifically $m_0^2 = \frac{1}{2}(m_{\tilde{b}}^2 + m_{\widetilde{Q}}^2)$), which allow proper bottom-tau unification for large $\tan\beta$. The shaded and hatched regions correspond to exact and approximate Yukawa-unified scenarios, respectively, while the solid and dashed lines represent the lower bounds imposed by LEP chargino limits assuming m_0 is 600 (the preferred value) or 200 GeV, respectively.

normalized to a typical squark mass m_0 (taken specifically to be $m_{\rm eff}$), for various values of the top quark mass, $\alpha_s(m_Z)$ and λ_t^G/λ_b^G. Also shown are the lower bounds on $m_{\widetilde{W}}/m_0$ and μ/m_0 imposed by LEP chargino limits, for $m_0 = 600\,\text{GeV}$ (solid lines) or $m_0 = 200\,\text{GeV}$ (dashed lines). Evidently the large, nonlogarithmic threshold correction δ_b is actually of great use: since, unlike the typical logarithmic corrections, it is very sensitive to the superspectrum, we can use experimental measurements of the top mass to learn about the hierarchies in the superspectrum. And from a model-building viewpoint, we can exclude those theories in which the gaugino mass and μ parameter do not fall into the allowed ranges shown in Fig. 3.

To reemphasize the importance of the corrections to the bottom mass (see also [25], we

have studied the consequences of bottom-tau unification for arbitrary λ_t^G/λ_b^G, imposing only that $\lambda_G \equiv \lambda_t^G < 2$ to ensure the validity of a perturbative expansion. (A larger λ_t^G would result in a weak-scale top mass very close to the fixed-point value, regardless of bottom-tau unification. This fixed-point value depends on λ_b^G and λ_τ^G, as well as on α_s.) In Fig. 4 we show the predictions for $\tan\beta$ and $m_t^{\rm pole}$, fixing the values of $\alpha_s(m_Z)$ and $m_b(m_b)$ and integrating over all possible superpartner and pseudoscalar Higgs masses between m_Z and 1 TeV. The black regions correspond to $\delta_b = 0$, while in the gray regions $-25\% < \delta_b < 20\%$. The effects of finite δ_b are striking.

3 $b \to s\gamma$

There is another immediate phenomenological implication of the large 1-loop corrections in the $\tan\beta \gg 1$ framework [15, 27]: the same diagrams, but with a photon attached in all possible ways and with a flavor-changing vertex as shown in Fig. 5, contribute to the bottom quark radiative decay $b \to s\gamma$. These contributions, which for small $\tan\beta$ are typically somewhat smaller than or comparable to the 2-Higgs standard model contribution, are parametrically enhanced by a factor of $\tan\beta \sim 50$ in the amplitude. But the CLEO bound [28] on the inclusive branching ratio $\mathrm{BR}(b \to s\gamma) < 4.2 \times 10^{-4}$ (at 95% CL) is already roughly saturated by the 2-Higgs standard model amplitude even if the charged Higgs is rather heavy (and is in fact oversaturated with a light charged Higgs). Hence the large $\tan\beta$ contribution must not be too large.

For two reasons [29], we will focus our attention on the Yukawa-coupled ("primary") chargino-exchange diagram of Fig. 5(a) rather than the gluino-exchange diagram of Fig. 5(b) or the gauge-coupled ("secondary") chargino-exchange diagram of Fig. 5(c). First, throughout the relevant regions of parameter space, the primary chargino-exchange amplitude exceeds or approximately equals the other amplitudes in magnitude. Second, it is predominantly determined by the third generation, namely by stop exchange, and hence its magnitude is fixed by the Kobayashi-Maskawa quark-mixing matrix element $V_{ts} \simeq V_{cb}$; in contrast, the other two diagrams arise from squark mixing between the second and third generation, and therefore depend on an independent mixing angle, which we shall call V_{23}. The primary chargino diagram also depends on the A parameter which mixes the SU(2)-doublet and -singlet stop squarks, but as we noted above, the RG equations typically fix A at low energies almost entirely in terms of the gaugino mass (independent of the GUT-scale value of A). So the sign and magnitude of this diagram is completely calculable [14, 16] in terms of the same parameters which enter δ_b; we have found that, when $\delta_b < 0$, as must be the case for any sizeable $|\delta_b|$, the chargino exchange amplitude $\mathcal{A}_{\widetilde{\chi}^-}$ interferes *constructively* with the standard model and charged-Higgs amplitudes ($\mathcal{A}_{\rm SM}$ and $\mathcal{A}_{H^-}$, respectively). Hence there can be no cancellations between these, and the constraint on $\mathcal{A}_{\widetilde{\chi}^-}$ is more severe. On the other hand, the new angle V_{23}, which determines the gluino and secondary chargino amplitudes $\mathcal{A}_{\widetilde{g}}$ and $\mathcal{A}'_{\widetilde{\chi}^-}$, gets not only a contribution $\sim V_{cb}$ through the RG evolution, but also one from the flavor structure at the GUT scale. Since we would like to remain as model-independent as possible, we will not make any assumptions about this structure, and thus V_{23} will not be

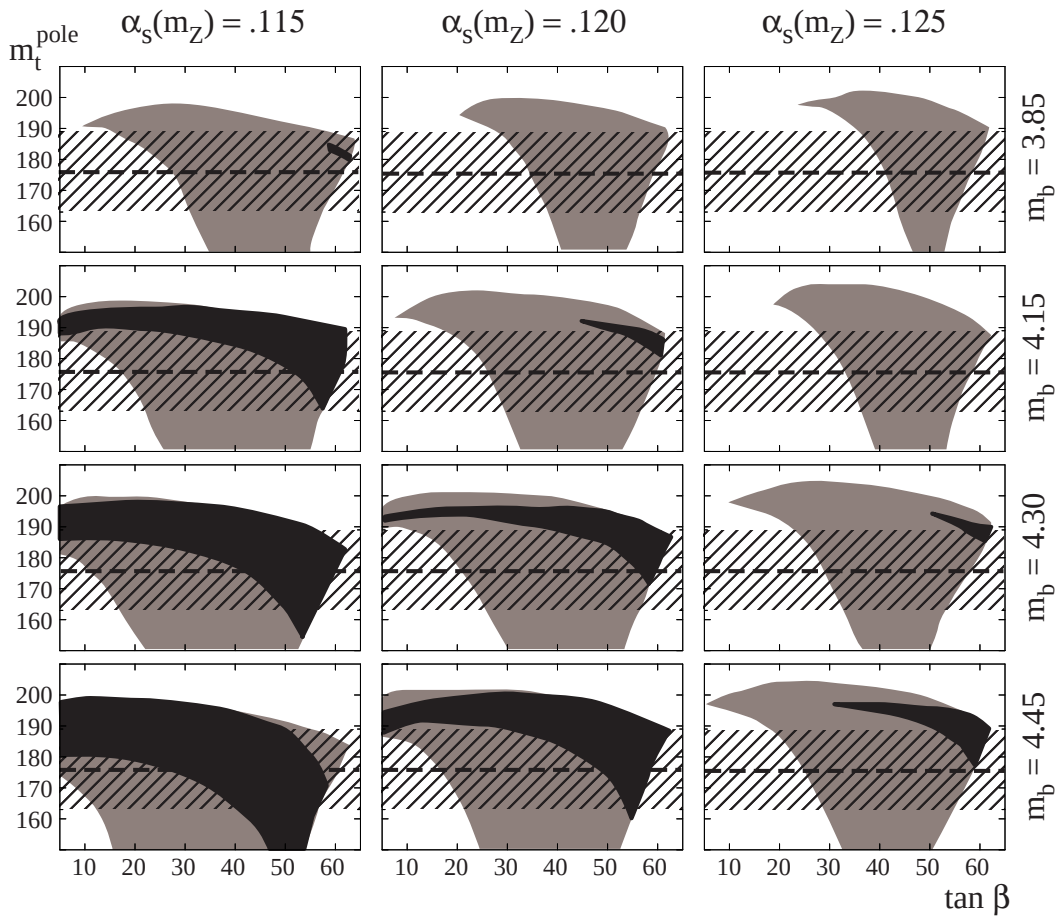

Fig. 4. The ranges of top quark pole mass and $\tan\beta$ allowed by bottom-tau unification at $M_{\rm GUT}$, for different values of the strong coupling and of the running bottom quark mass. We scan over all possible superspectra between m_Z and 1 TeV, allow arbitrary λ_t^G/λ_b^G, and restrict $\lambda_G \equiv \lambda_t^G < 2$ to ensure perturbativity. The effects of the 1-loop corrections δ_b are evident by comparing the black regions, for which $\delta_b = 0$, with the gray ones, in which $-25\% < \delta_b < 20\%$. These predictions are expected to be accurate to within a few GeV. The dashed horizontal line and hatched band are the top mass central value and an estimate of its uncertainty, respectively, as recently announced by the CDF collaboration [26].

determined. By tuning the flavor parameters and thereby V_{23} one could cancel the various large $\tan\beta$ contributions to $b \to s\gamma$ against each other and avoid any bounds from this process. However, one should also be careful about other phenomenological implications of this new source of flavor violation. Since we are dealing with a grand unified theory, there is also a leptonic counterpart of the new mixing angle V_{23}, and this gives relevant contributions to the rate for $\tau \to \mu\gamma$. A more detailed analysis of the potential for cancellations and of $\Gamma(\tau \to \mu\gamma)$ is presented elsewhere [29]. Our approach here will be to take into account only the sizeable

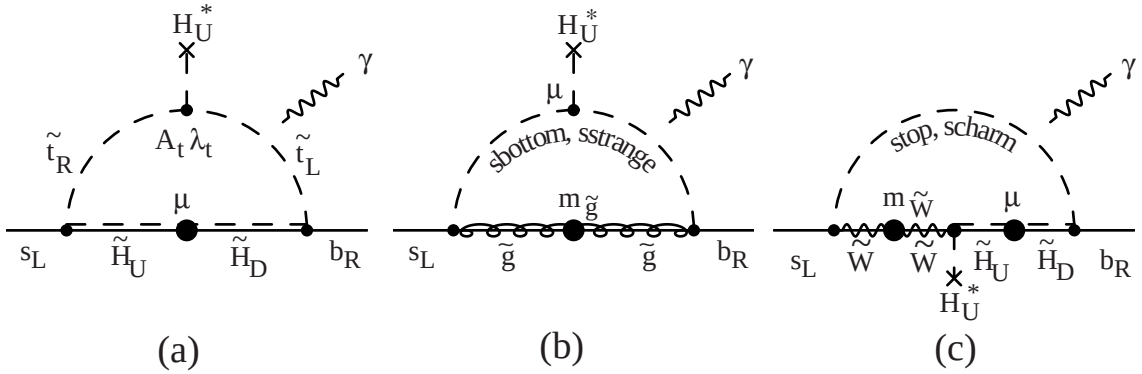

Fig. 5. The primary chargino, gluino, and secondary chargino diagrams which contribute to radiative bottom quark decay and are enhanced by large $\tan\beta$. Note that while the amplitude in (a) is completely predicted by third-generation parameters and V_{cb}, the other two depend sensitively on intragenerational squark mixing parameters and may be equally important.

and calculable primary chargino contribution: in those parameter ranges for which it is small, the other amplitudes are typically also small and there is no conflict with experiment, while in those ranges for which $\mathcal{A}_{\rm SM} + \mathcal{A}_{H^-} + \mathcal{A}_{\widetilde{\chi}^-}$ exceeds the experimental bounds, the other diagrams must be tuned to cancel against these amplitudes. Our goal is to study when and to what degree such a cancellation is needed.

We will use the following expressions [30] to calculate the limits on the MSSM parameters for large $\tan\beta$:

$$\mathrm{BR}(b\to s\gamma) \simeq \mathrm{BR}(b\to ce\bar{\nu})\frac{(6\alpha_{\rm em}/\pi)\left[\eta^{16/23}A_\gamma + C\right]^2}{I(m_c/m_b)\left[1-(2/3\pi)\alpha_s(m_b)f_{\rm QCD}(m_c/m_b)\right]} \tag{6}$$

where $\mathrm{BR}(b\to ce\bar{\nu}) = 0.107$, $\eta = \alpha_s(m_Z)/\alpha_s(m_b) = 0.520$ [using $\alpha_s(m_Z) = 0.12$ as a reference value], $C = -0.191$, $I(m_c/m_b) = 0.45$, $f_{\rm QCD}(m_c/m_b) = 2.41$. The dimensionless amplitude is given by $A_\gamma = \mathcal{A}_{\rm SM} + \mathcal{A}_{H^-} + \mathcal{A}_{\widetilde{\chi}^-}$ where

$$\mathcal{A}_{\rm SM} = \frac{3}{2}\frac{m_t^2}{m_W^2}f_\gamma^{(1)}\left(\frac{m_t^2}{m_W^2}\right) \tag{7}$$

$$\mathcal{A}_{H^-} = \frac{1}{2}\frac{m_t^2}{m_{H^-}^2}f_\gamma^{(2)}\left(\frac{m_t^2}{m_{H^-}^2}\right) \tag{8}$$

$$\mathcal{A}_{\widetilde{\chi}^-} = \tan\beta\,\frac{m_t^2 A_t\mu}{m_{\tilde{t}}^4}\sum_{j=1}^{2}U_{j2}V_{j2}^*\frac{m_{\widetilde{\chi}_j^-}}{\mu}F\left(\frac{m_{\tilde{t}}^2}{m_{\widetilde{\chi}_j^-}^2}\right) \tag{9}$$

and the various functions are

$$f_\gamma^{(1)}(x) = \frac{7-5x-8x^2}{36(x-1)^3} + \frac{x(3x-2)}{6(x-1)^4}\ln x \tag{10}$$

$$f_\gamma^{(2)}(x) = \frac{3-5x}{6(x-1)^2} + \frac{3x-2}{3(x-1)^3}\ln x \tag{11}$$

$$F(x) = \frac{x}{6(1-x)^3}\left[5-12x+7x^2+2x(2-3x)\ln x\right] . \tag{12}$$

In Eq. (9) we assume that the two stop mass eigenvalues are roughly degenerate. (This is in particular a good approximation for the interesting situation in which the stop is heavier then the top and the diagonal soft masses are almost degenerate.) Notice that the crucial off-diagonal $\tilde{t}_R\tilde{t}_L$ mixing has been factored out in $\mathcal{A}_{\widetilde{\chi}^-}$. We have kept the exact dependence on the chargino mass matrix:

$$U^* \begin{pmatrix} m_{\widetilde{W}} & m_W\sqrt{2}\sin\beta \\ m_W\sqrt{2}\cos\beta & \mu \end{pmatrix} V^{-1} = \begin{pmatrix} m_{\widetilde{\chi}_1^-} & 0 \\ 0 & m_{\widetilde{\chi}_2^-} \end{pmatrix} . \tag{13}$$

We can now determine how large $\mathcal{A}_{\widetilde{\chi}^-}$ can be without requiring some cancellation to avoid conflict with experiment. We find that if $\mathcal{A}_{\widetilde{\chi}^-} = 0$ then the 2-Higgs standard model exactly saturates the experimental bound for $m_{H^-} = 1\,\mathrm{TeV}$ and $m_t = 170\,\mathrm{GeV}$; for those values $\mathcal{A}_{H^-} \simeq .15\mathcal{A}_{\mathrm{SM}}$. Assuming some theoretical uncertainty allows the charged Higgs to be significantly lighter: a 30% uncertainty in the branching ratio would allow a charged Higgs of 350 GeV ($\mathcal{A}_{H^-} \simeq .5\mathcal{A}_{\mathrm{SM}}$), while with a 50% uncertainty the charged Higgs can be as low as 250 GeV ($\mathcal{A}_{H^-} \simeq .75\mathcal{A}_{\mathrm{SM}}$). For the ranges of top quark mass we are considering, the m_t dependence is much smaller than such theoretical uncertainties. Now, if we add constructively a supersymmetric contribution equal to 50% of the standard-model amplitude, the branching ratio is 30% above the experimental bound without including any charged-Higgs amplitude, or 50% above the experimental bound if we include a 25% charged-Higgs contribution. (These values correspond to taking the soft SUSY-breaking scalar masses to be roughly degenerate at $\sim 700\,\mathrm{GeV}$ while the μ and wino mass are at $\sim m_Z$, which we argue is the favored scenario.) The charged Higgs mass cannot be made too large without fine-tuning the Z mass, so a 25% charged-Higgs contribution is a reasonable lower bound. Thus, to be conservative, we conclude that *there must be some cancellations whenever* $\mathcal{A}_{\widetilde{\chi}^-} > .5\mathcal{A}_{\mathrm{SM}}$, either from the other diagrams [Figs. 5(b) and 5(c)] due to tuning of the flavor physics or from cancellations in the A parameter due to tuning of the GUT values of A and the gaugino mass. We define a measure of the need for cancellations as[4]:

$$\epsilon_{b\to s\gamma} \equiv \min\left(\left|\frac{.5\mathcal{A}_{\mathrm{SM}}}{\mathcal{A}_{\widetilde{\chi}^-}}\right|, 1\right) . \tag{14}$$

[4]Since $\mathcal{A}_{\widetilde{\chi}^-} = .5\mathcal{A}_{\mathrm{SM}}$ requires no more tuning than $\mathcal{A}_{\widetilde{\chi}^-} < .5\mathcal{A}_{\mathrm{SM}}$, we define $\epsilon_{b\to s\gamma}$ so it saturates when $\mathcal{A}_{\widetilde{\chi}^-} = .5\mathcal{A}_{\mathrm{SM}}$. For similar reasons we will define $\epsilon_B \equiv \min(B/m_{\widetilde{W}}, 1)$ and $\epsilon_B \equiv \min(m_Z^2/m_A^2, 1)$.

For example, if the chargino amplitude is 10 times greater than the saturating amplitude $.5\mathcal{A}_{\rm SM}$, then some other contribution must be adjusted with a precision of 0.1 to cancel against it. In fact, we will never need the exact definition of $\epsilon_{b\to s\gamma}$; all that we will require of it is that it be small whenever the chargino amplitude is considerably too big.

4 Electroweak Symmetry Breaking

Thus far, the existence of the electroweak-breaking Higgs VEVs v_U and v_D has been simply assumed. We must study the generation of those VEVs in order to understand how compatible is the idea of Yukawa unification with radiative electroweak breaking, how the SO(10) or SU(5) symmetry constraints on the soft SUSY-breaking mass parameters affect this breaking, how the top-bottom mass hierarchy is obtained, and how natural or unnatural is such a scenario. This study will reveal favored ranges for the Yukawa couplings and the soft SUSY-breaking masses, and highlight the central role of the PQ and R symmetries in discussing electroweak breaking for Yukawa-unified models.

Electroweak symmetry breaking is governed by the scalar potential for the neutral components of the two Higgs fields H_U and H_D, which at tree level is of the form

$$\begin{aligned} V_0 &= m_U^2|H_U|^2 + m_D^2|H_D|^2 + \mu B\,(H_U H_D + \text{h.c.}) + \\ &\quad \frac{g_1^2+g_2^2}{8}\left(|H_U|^2 - |H_D|^2\right)^2 \end{aligned} \tag{15}$$

where g_2 and g_1 are respectively the $SU(2)$ and hypercharge gauge couplings. The parameters $m_{U,D}^2 = \mu_{U,D}^2 + \mu^2$ contain the soft-breaking Higgs masses as well as the μ parameter from the superpotential, while B is the soft-breaking bilinear mass parameter. The conditions for proper electroweak breaking are well-known:

$$m_U^2 + m_D^2 \geq 2|\mu B| \tag{16}$$

ensures that the potential is bounded from below, and

$$m_U^2 m_D^2 < \mu^2 B^2 \tag{17}$$

guarantees the existence of a minimum away from the origin and so breaks the symmetry. In practice, since $|\mu B|$ will always be much less than or at most comparable to $|m_U^2|$ and $|m_D^2|$, we can reduce these requirements to $m_A^2 = m_U^2 + m_D^2 > 0$ (using the expression for the pseudoscalar Higgs mass) and $m_U^2 < 0$ (noting that large $\tan\beta$ means that the up-type Higgs gets the large VEV).

It is useful to interpret the two above conditions for proper symmetry breaking in terms of the usual custodial symmetry exchanging up- and down-type particles. In practice we need $m_U^2 < 0 < m_D^2$, which represents a substantial violation of this symmetry. In the usual scenario, with the initial condition $\lambda_t \gg \lambda_{b,\tau}$, this custodial violation is provided by the Yukawa couplings themselves. The large top Yukawa coupling drives negative the soft-breaking mass

parameter of the "up-type" Higgs H_U to which it couples, while the other Higgs is largely unaffected. Hence, even with custodially symmetric soft masses at M_G, it is very natural to obtain [11, 12], at the electroweak scale, $m_U^2 < 0$ while the other Higgs mass is sufficiently positive to ensure $m_U^2 + m_D^2 > 0$. On the other hand, with unified or approximately-unified Yukawa couplings ($\lambda_t^G/\lambda_b^G \sim 1$), it is hard to see why the two Higgs mass parameters should run differently, so the natural explanation for $m_U^2 < 0 < m_U^2 + m_D^2$ is lost. In fact, with the boundary condition $\lambda_t^G = \lambda_b^G$, the only sources of custodial breaking in the couplings are the hypercharges and the absence of a right-handed neutrino (but presence of a right-handed tau). These furnish only a tiny splitting even when integrating from the GUT to the electroweak scales. If such a splitting between m_U^2 and m_D^2 is to be $\sim m_Z^2$ then the soft-breaking masses themselves must be considerably bigger than m_Z^2. To break the symmetry more naturally, custodial breaking must be enhanced. To this end, either m_U^2 can be split by various means from m_D^2 already at the GUT scale, or one may relax the requirement $\lambda_t^G = \lambda_b^G$, which in any case is bound to be modified—either by a little, due to GUT-scale threshold corrections, or by a lot, in the case of mixed-SO(10) or SU(5) models.

Furthermore, in the usual scenario, if all mass parameters in the scalar potential are comparable then so are the VEVs of the two Higgs doublets; but comparable Higgs VEVs are perfectly compatible in the usual scenario with the experimental hierarchy $m_t \gg m_b$ since that is furnished by the assumed hierarchy of Yukawa couplings $\lambda_t \gg \lambda_b$. By contrast, in a unified scenario it is the ratio of VEVs which must be large. Upon minimizing the scalar potential V_0, one easily obtains (for large $\tan\beta$)

$$\frac{1}{\tan\beta} \equiv \frac{v_D}{v_U} = -\frac{\mu B}{m_u^2 + m_D^2} = -\frac{\mu B}{m_A^2}\,, \tag{18}$$

as well as $m_Z^2 = -2m_U^2$ (which sets the scale). A large VEV hierarchy requires a small coupling between the two Higgs doublets, namely $\mu B \ll m_U^2 + m_D^2 = m_A^2$, so that an expectation value in one is only weakly fed into the other. But this small Higgs coupling, as we now discuss, implies [21, 16] a necessary degree of cancellation between some parameters at the GUT scale.

5 Generating a Hierarchy

From Eq. (18) it is clear that to generate the hierarchy of VEVs ($\tan\beta \sim 50$) necessary for the top-bottom (and top-tau) mass ratio, we need a small μB or a large m_A^2. This is difficult: on one hand, $m_A^2 = m_U^2 + m_D^2$ and $m_Z^2 = -2m_U^2$ run quite similarly and are tightly coupled through the RG evolution, so it is difficult to make m_A^2 much larger than the electroweak scale; on the other hand, μ cannot be much below the Z mass since otherwise a light chargino (or neutralino) would have been detected at LEP, and the RG equations imply that B is naturally at least as large as $m_{\widetilde{W}}$—which again cannot be much below the Z mass without producing a

light chargino or neutralino. To make these arguments concrete, we can rewrite Eq. (18) as

$$\frac{1}{50} \sim -\frac{\mu\, B}{m_A^2} = -\underbrace{\frac{\mu\, m_{\widetilde{W}}}{m_Z^2}}_{\gtrsim 0.9} \cdot \underbrace{\frac{B}{m_{\widetilde{W}}}}_{\equiv \epsilon_B} \cdot \underbrace{\frac{m_Z^2}{m_A^2}}_{\equiv \epsilon_Z} \ . \tag{19}$$

The most natural scenario would display approximate PQ or R symmetries, with either $\mu \ll m_Z$ or $m_{\widetilde{W}} \sim B \sim A \ll m_Z$ or both, while all other masses would be around m_Z. When $\tan\beta \gg 1$, such small μ or $m_{\widetilde{W}}$ result in a chargino much lighter than m_Z. Alas, this scenario is now experimentally ruled out: specifically, LEP [31] places a lower bound of roughly $\frac{1}{2}\, m_Z$ on the lightest chargino mass. A lower bound of $\frac{1}{2}\, m_Z$ on the lightest eigenvalue of the chargino mass matrix in Eq. (13) translates into a bound on the Lagrangian parameters of

$$\left[m_{\widetilde{W}}^2 - \left(\tfrac{1}{2}\, m_Z\right)^2\right] \left[\mu^2 - \left(\tfrac{1}{2}\, m_Z\right)^2\right] \gtrsim 2\, m_W^2 \left(\tfrac{1}{2}\, m_Z\right)^2 . \tag{20}$$

Subject to this bound, the prefactor $\mu m_{\widetilde{W}}/m_Z^2$ in Eq. (19) is minimized when $\mu = m_{\widetilde{W}}$, yielding $\mu m_{\widetilde{W}}/m_Z^2 \gtrsim 1/4 + 2^{-1/2} m_W/m_Z \simeq .88$. Therefore, the parameters at the GUT scale must be adjusted so that, at the electroweak scale, either B is much lighter than its natural minimum value $\sim m_{\widetilde{W}}$, or m_Z^2 is much less than its natural value $\sim m_A^2$. (Note that in the first case $\tan\beta$ is the quantity that is tuned to be small, while in the second case m_Z^2 is tuned to be small.) We quantify these tunings by ϵ_B and ϵ_Z, respectively [see the footnote for Eq. (14)]. Thus the bounds from LEP imply that some GUT parameters must be adjusted to cancel with an accuracy of at least

$$\epsilon_B \epsilon_Z \ \lesssim\ \frac{1}{\tan\beta} \sim \frac{1}{50}\,; \quad \text{saturated when}\, \mu \sim m_{\widetilde{W}} \sim m_Z \tag{21}$$

and the least amount of tuning is required when μ and $m_{\widetilde{W}}$ are roughly as light as they can be, namely both near the Z mass.

Eq. (19) by itself does not distinguish whether B or m_Z^2 should be tuned small; that is decided by $\epsilon_{b\to s\gamma}$, which requires considerable tuning of the flavor sector or of the A parameter at the GUT scale if the superpartners and the pseudoscalar Higgs are near m_Z. Table 2 sketches four possible scenarios, distinguished by whether PQ and R are good symmetries (at low energies—m_i are the sfermion masses at the electroweak scale). Note that $\epsilon_B \epsilon_Z$ favor the PQ- and R-symmetric case *or* the asymmetric case having all masses near m_Z, but that $\epsilon_{b\to s\gamma}$ favors the former over the latter. Thus the most natural scenario as measured by these three criteria is the maximally symmetric one: a Lagrangian ([16]; see also the first reference in [32]) which is maximally PQ- and R-symmetric while obeying the LEP bounds and keeping the superpartners as light as possible, that is,

$$\begin{cases} m_A \sim m_0 \sim \sqrt{\tan\beta}\, m_Z \ (\sim 600\,\text{GeV}) \\ \mu \sim A \sim B \sim m_{\widetilde{W}} \sim \frac{1}{3} m_{\tilde{g}} \sim m_Z \end{cases} \tag{22}$$

where by m_0 we mean the typical mass of the squarks and sleptons evaluated at the electroweak scale. (To reiterate: we chose $\mu \sim m_{\widetilde{W}} \sim \frac{1}{7} m_0$ because this is the most natural case allowed by

	PQ	no PQ
no R	$\mu \sim \frac{1}{7} m_0$ $m_{\widetilde{W}} \sim m_0$ $m_Z \sim \frac{1}{7} m_0 \implies \epsilon_Z \sim \frac{1}{50}$ $m_i \sim m_0 \implies \epsilon_{b\to s\gamma} \sim 1$ $\epsilon_B \sim \frac{1}{7}$	$\mu \sim m_0$ $m_{\widetilde{W}} \sim m_0$ **either** $m_Z \sim \frac{1}{7} m_0 \implies \epsilon_Z \sim \frac{1}{50}$ $m_i \sim m_0 \implies \epsilon_{b\to s\gamma} \sim 1$ $\epsilon_B \sim \frac{1}{50}$ **or** $m_Z \sim m_0 \implies \epsilon_Z \sim 1$ $m_i \sim m_0 \implies \epsilon_{b\to s\gamma} \sim \frac{1}{50}$ $\epsilon_B \sim \frac{1}{50}$
R	$\mu \sim \frac{1}{7} m_0$ $m_{\widetilde{W}} \sim \frac{1}{7} m_0$ $m_Z \sim \frac{1}{7} m_0 \implies \epsilon_Z \sim \frac{1}{50}$ $m_i \sim m_0 \implies \epsilon_{b\to s\gamma} \sim 1$ $\epsilon_B \sim 1$	$\mu \sim m_0$ $m_{\widetilde{W}} \sim \frac{1}{7} m_0$ $m_Z \sim \frac{1}{7} m_0 \implies \epsilon_Z \sim \frac{1}{50}$ $m_i \sim m_0 \implies \epsilon_{b\to s\gamma} \sim 1$ $\epsilon_B \sim \frac{1}{7}$

Table 2: Typical scenarios, and consequent fine-tunings, for generating the Higgs VEV hierarchy $v_D \sim 1/50 v_U$ with or without the PQ and R symmetries.

LEP—if μ or $m_{\widetilde{W}}$ were smaller, m_Z would require further tuning to make it sufficiently light, while if they were much larger we would lose the advantages of the PQ and R symmetries and the tuning would again be exacerbated.) This spectrum implies a small correction to the bottom mass, $\delta_b \sim 5\%$, and hence (from Table 1) a heavy top and preferably a somewhat low value of $\alpha_s(m_Z)$.

6 Correct Symmetry Breaking

6.1 The General Problem

We return now to the question of how the electroweak symmetry may be correctly broken, while preserving the $\mathrm{SU(3)_c} \times \mathrm{U(1)_{em}}$ gauge symmetries. In principle, what needs to be done is to study the effective potential $V_{\rm eff}$ for field values $m_S \leq \phi \leq M_G$, where m_S denotes collectively the soft SUSY-breaking masses and ϕ is the set of scalar fields in the MSSM. In practice, since we are dealing with a perturbative theory, we need only consider the RG-improved tree level potential $V_0(\Lambda)$ renormalized at scales Λ between m_S and $M_{\rm GUT}$. First, we must make sure that the potential is stable for $\Lambda \gg m_S$, i.e. that $V_0(\Lambda)$ is bounded from below at high scales. If this were not the case, the scale of gauge symmetry breaking would be $\sim \Lambda \gg m_S$, which is phenomenologically unacceptable, and possibly even $\mathrm{SU(3)_c} \times \mathrm{U(1)_{em}}$ would be broken. Second, we need to guarantee that at scales $\Lambda \sim m_S$ the minimum of $V_0(\Lambda)$

is such that H_u and H_d, and no other fields, acquire nonvanishing expectation values. This amounts to imposing some positivity constraints on the sfermion mass-squared parameters, which we discuss below. Finally, since we are interested in large rather than small $\tan\beta$, the instability should arise in the H_u direction, while the small H_d VEV is generated through the mixing mass parameter $B\mu$.

An essentially technical comment is in order here. We will be mainly concerned with the parameters related to the third family and to the Higgs sector. This is because, in the limit in which flavor mixings and the Yukawa couplings of the two light families are neglected, the SUSY-breaking masses of these families have a numerically small impact on the RG evolution of the parameters of the third generation and the Higgs sector. The only effect [33] is via the hypercharge D-term S (see Appendix A), which is small because of the small hypercharge gauge coupling. Moreover the effects of S are completely determined by its GUT scale value S_G, since it renormalizes multiplicatively. Indeed with SU(5) or SO(10) boundary conditions on scalar masses, S_G itself is completely specified by the soft masses of H_u and H_d (since these are the only light incomplete matter SU(5) multiplets).

The relevant parameters then consist of the seven soft-breaking scalar masses $\mu_U^2 = m_U^2 - \mu^2$, $\mu_D^2 = m_D^2 - \mu^2$, $m_{\tilde{t}}^2$, $m_{\tilde{b}}^2$, $m_{\widetilde{Q}}^2$, $m_{\tilde{\tau}}^2$ and $m_{\widetilde{L}}^2$, the three trilinear soft-breaking parameters $A_{t,b,\tau}$, the single bilinear soft-breaking parameter B, the μ term in the superpotential, and the three gaugino masses. Their 1-loop RG equations are given for reference in Appendix A.

Let us now discuss in more detail the constraints which these parameters need to satisfy. We begin with those which must be satisfied at scales $\Lambda \gg m_S$. It is well known that the MSSM, like any generic SUSY model, possesses a host of "approximately flat" directions in the space of scalar fields ϕ. By "approximately flat" we mean that the potential, at the renormalizable level, is only quadratic in those directions. In general, though, irrelevant operators suppressed by inverse powers of a large mass M such as $M_{\rm GUT}$ or $M_{\rm Planck}$ can give an additional stabilizing contribution [34]. To be conservative, we will always assume the superpotential contains an appropriate operator of the form ϕ^4/M. Then along any such direction parametrized by a field ϕ_α, the potential is essentially (see below) given by $m_\alpha^2|\phi_\alpha|^2 + |\phi_\alpha|^6/M^2$, where m_α^2 is equal to a sum of squared masses. Regardless of the sign of m_α^2, there is no minimum for $\phi_\alpha \gtrsim \sqrt{m_\alpha M} \equiv \Lambda_{\rm HIGH}$ ($\sim 10^9\,{\rm GeV}$ for $M = M_{\rm GUT}$), so parameters normalized at scales $\Lambda > \Lambda_{\rm HIGH}$ can never yield an unwanted minimum. For $\phi_\alpha < \Lambda_{\rm HIGH}$, the potential in dominated by the quadratic piece, though there may be a scale $\Lambda_{\rm LOW}$ below which a linear term may again stabilize the potential. In the absence of a linear term, the lowest scale of interest is $\Lambda_{\rm LOW} \sim m_S$, at which the superpartners are integrated out. If $m_\alpha^2 > 0$ were to become negative at a critical scale $\Lambda_{\rm LOW} \ll \Lambda_c \ll \Lambda_{\rm HIGH}$, dimensional transmutation [35] would take place: the VEV of ϕ_α would be fixed by the one loop correction to the effective potential $V_1(\Lambda)$ to be of order Λ_c (times a coupling constant). (Notice that, in the absence of the irrelevant operator, if m_α^2 were to be negative already at the GUT scale, then we would clearly be expanding around the wrong vacuum in the GUT theory.) In order to get acceptable low-energy physics we have then to impose $m_\alpha^2(\Lambda) \geq 0$, for all approximately-flat directions α and for all scales Λ between the $\Lambda_{\rm HIGH}$ and $\Lambda_{\rm LOW}$ relevant to that ϕ_α.

When we restrict our attention to the fields of the third family and the Higgs sector, there

are only two such flat directions:

1. $\langle H_u \rangle = \langle H_d \rangle = \phi_1$, with all other fields at zero; and
2. $\langle H_u \rangle = \phi_2$, $\langle \widetilde{L} \rangle = (\phi_2^2 + \phi_2 \mu/\lambda_b)^{1/2}$, $\langle \widetilde{Q} \rangle = \langle \widetilde{b_c} \rangle = (\phi_2 \mu/\lambda_b)^{1/2}$, with all other fields at zero [36].

The color and isospin orientations are determined by imposing vanishing D and F terms. Along ϕ_1 the potential is purely quadratic, $V_0(\phi_1) = m_1 |\phi_1|^2$ with $m_1^2 = m_U^2 + m_D^2 - 2|B\mu|$. The stability constraint has already been given in Eq. (16). This constraint should be satisfied between $\Lambda_{\rm HIGH} \sim 10^9\,{\rm GeV}$ (to be conservative) and $\Lambda_{\rm LOW} \sim m_S$. Along direction ϕ_2 there is also a linear term:

$$\begin{aligned} V_0(\phi_2) &= (\mu_U^2 + m_{\widetilde{L}}^2)|\phi_2|^2 + (m_{\widetilde{L}}^2 + m_{\widetilde{Q}}^2 + m_{\widetilde{b}}^2) \left| \frac{\mu \phi_2}{\lambda_b} \right| \\ &\equiv m_2^2 |\phi_2|^2 + m_3^2 \left| \frac{\mu \phi_2}{\lambda_b} \right| . \end{aligned} \tag{23}$$

For this flat direction, the dominant stabilizing term at high scales is the left-handed neutrino mass operator $(H_u L)^2/M_N$, where M_N is the right-handed neutrino mass. Indeed the effect of this operator can be important down to $\Lambda_{\rm HIGH} \sim 10^7\,{\rm GeV}$, since M_N could be as low as $10^{12} - 10^{13}\,{\rm GeV}$. At low scales, the linear term will stabilize the potential (provided $m_3^2 > 0$, which will always be the case). Therefore, as we show in Appendix B (see also Ref. [37]), the ϕ_2 flat direction can only pose a problem at scales above $\Lambda_{\rm LOW} \sim (2\pi/\alpha)\mu/\lambda_b \sim 10^4 - 10^5\,{\rm GeV}$. So we need to impose $m_2^2(\Lambda) > 0$ at least for all Λ between $\Lambda_{\rm HIGH} \sim 10^7\,{\rm GeV}$ and $\Lambda_{\rm LOW} \sim 10^5\,{\rm GeV}$.

A general scan of the values of $m_{1,2}^2(\Lambda)$ at all intermediate scales would be numerically arduous. Fortunately, with only minor assumptions on the initial parameters, $m_1^2(\Lambda)$ and $m_2^2(\Lambda)$ decrease essentially monotonically with Λ. Imposing positivity just at low Λ then guarantees the absence of unwanted vacua at all scales. Consider for instance the PQ- and R-symmetric limit of the RG equations in Appendix A. Monotonicity of $m_{1,2}^2$ is clearly satisfied when $X_{t,b,\tau}$ are positive throughout the running. In turn this condition is satisfied when the X_i start out positive and of comparable magnitudes (check for instance the entries in the matrix $\mathcal{H}$ in Appendix A, whose behavior in monotonic in Λ). In most interesting cases, the necessary positivity of masses at low energy will imply positive X_i at the GUT scale [for instance in minimal SO(10)]. Introducing a finite μ does not alter the conclusions, as long as R symmetry is preserved. For small μ and large gaugino masses, the situation is also unchanged: in the first stage of the running, their contribution to $m_{1,2}^2$ is positive, but very small; however it soon becomes negative and its absolute value increases monotonically when Λ is lowered, so again checking positivity of $m_{1,2}^2$ at low scales suffices. Finally, when both PQ and R are broken, the above discussion applies straightforwardly to m_2^2, but not to m_1^2 due to the additional inhomogeneous piece $B\mu$. For this situation we have explicitly verified monotonicity for a wide range of initial parameters. We thus conclude that, quite generally, the imposition of the constraints at a low scale is sufficient to ensure their validity

throughout the RG evolution. Our analysis is thereby considerably simplified: we need only impose $m_1^2(\Lambda = m_S \sim m_Z) \simeq m_A^2 > 0$ and

$$m_2^2(\Lambda \sim 10^5\,\mathrm{GeV}) > 0 \tag{24}$$

to avoid an instability in the ϕ_1 and ϕ_2 directions.

We note in passing that the constraints from flat directions involving also the fields in the first two families are not a problem for us. This is because their m_α^2 always involve the soft masses for these fields, which are for us essentially arbitrary and can thus be taken large enough to stabilize the flat directions.

We next turn to the constraints on the potential at the electroweak scale. In what follows we will use just the tree level potential $V_0(\Lambda)$. This approximation, discussed in detail in Appendix C, is motivated by the fact that we are not concerned with precise predictions for the various masses, but rather with the characteristic hierarchies in the spectrum, with the rough bounds on the various parameters and with comparing the naturalness of various parameter choices.

First, the scalar configuration with $\langle H_{u,d}\rangle \neq 0$, and all the other fields at zero, should be a local minimum. This will be the case if we impose that the MSSM parameters, evaluated at the electroweak scale, satisfy:

$$m_i^2(\Lambda = m_Z) > 0, \quad i = Z, A, \tilde{t}, \tilde{b}, \widetilde{Q}, \tilde{\tau}, \widetilde{L} \tag{25}$$

where as before $m_Z^2 = -2m_U^2$ and $m_A^2 = m_U^2 + m_D^2$. (Notice that in the above equation we have neglected any contribution to the sfermion masses coming from the Higgs VEV. We have also ignored the phenomenological bounds on these masses, which yield somewhat stronger constraints: $m_i^2 \gtrsim m_Z^2$. However we stress once more that, for the purpose of studying the spectrum hierarchies and the naturalness of different scenarios, the above contraints are sufficient. Indeed, in most situations we will end up with sfermion masses well above m_Z.)

A second class of constraints is needed to avoid having other minima with electroweak- or color-breaking VEVs of order m_S. Such minima can arise, even for positive sfermion masses, from the destabilizing effect of the trilinear terms in the scalar potential. These are given by the soft A-terms and also by the trilinear terms proportional to μ in the supersymmetric part of the scalar potential. In what follows we will mainly be concerned with necessary constraints, and will not enter into a comprehensive discussion of the sufficient ones. Let us consider the effect of A terms first. These were discussed in Ref. [38] where a necessary condition to avoid unwanted minima was given: $m_a^2+m_b^2+m_c^2 \geq |A|^2/3$ (where a, b, c represent any three fields having a Yukawa coupling λ, and A is the corresponding soft-SUSY-breaking trilinear coupling). When this condition is not satisfied, there is a color- and charge-breaking minimum with energy density $\sim -m_S^4/\lambda^2$. In the case of a light fermion this vacuum is considerably deeper than the usual Higgs one. For the top quark, λ is sufficiently large, and m_S is often assumed to be sufficiently small, that this extraneous vacuum is not deeper (and typically shallower) than the ordinary vacuum. This is why the A-term requirement is usually not applied to the stop. However, when there is a hierarchy $m_S \gg m_Z$, the

extraneous minimum, when it exists, is indeed parametrically deeper than the $\mathcal{O}(g^2v^4)$ Higgs minimum, so the necessary condition given above must be applied also to the soft parameters of the third generation. Similar arguments can be made for the trilinear μ terms, though we are not aware of previous discussions in the literature. Now the triplets of fields in danger of developing expectation values are those entering the various μ couplings: $(H_u, \widetilde{L}, \widetilde{\tau})$, $(H_u, \widetilde{Q}, \widetilde{b})$ and $(H_d, \widetilde{Q}, \widetilde{t})$ (where the last member of each triplet is the SU(2)-singlet scalar field). For instance, along the direction $\langle H_u \rangle = \langle \widetilde{L} \rangle = \langle \widetilde{\tau} \rangle = \phi$ the potential is given by

$$V = (m_U^2 + m_{\widetilde{L}}^2 + m_{\widetilde{\tau}}^2)|\phi|^2 - 2\mu\lambda_\tau|\phi|^3 + \hat{\lambda}_\tau^2|\phi|^4 \tag{26}$$

where $\hat{\lambda}_\tau^2 = \lambda_\tau^2 + (g_1^2 + g_2^2)/2$. To avoid a minimum away from the origin in the above potential, μ must not be too big[5]:

$$\lambda_\tau^2\mu^2 < \hat{\lambda}_\tau^2 \left(m_U^2 + m_{\widetilde{L}}^2 + m_{\widetilde{\tau}}^2 \right) . \tag{27}$$

Notice that, because of the D-term contribution to $\hat{\lambda}_\tau^2$, the bound (27) is irrelevant when the Yukawa couplings are small (namely for sfermions of the first two families, or even τ and b at small $\tan\beta$.) When μ is somewhat above this bound, an unwanted minimum with $V \sim -\mu^4\lambda_\tau^4/\hat{\lambda}_\tau^6$ is present. Again, for scenarios with a hierarchy $\mu \gg m_Z$, this new vacuum is much deeper than the correct one.

In the course of our study we have verified that the positivity constraints of Eqs. (24) and (25) are always stronger than those coming from the trilinear A and μ terms, at least *for the parameter ranges of interest to us*. Thus, while important in principle, the instabilities arising from trilinear terms in the scalar potential do not impose any constraints in practice.

Next, we examine the evolution of the Lagrangian parameters down to the electroweak scale. The form of the RG equations dictates that the soft-breaking scalar mass-squared parameters at the electroweak scale are linearly related to their initial values, to the square of the GUT-scale gaugino mass $M_{1/2}$, and to the square of the μ parameter. (In fact there are also terms proportional to the GUT-scale values of the A parameters, namely $\propto A_G^2$ and $\propto A_G M_{1/2}$. As shown in Appendix A they can be neglected unless A_G is at least an order of magnitude bigger than the other GUT-scale parameters.) Thus the constraints of Eqs. (25) and Eq. (24), when saturated, constitute a set of eight hyperplanes in the space of initial scalar mass-squared parameters. When $M_{1/2} = \mu = 0$, the low-energy masses m_i^2 are just homogeneous linear combinations of the GUT-scale parameters M_i^2, so the various constraint planes determine a cone—or rather, technically, a pyramidal surface—within which those constraints are satisfied. Such a cone, drawn in the 3-dimensional space of initial parameters for the minimal SO(10) theory as discussed below, is shown in Fig. 6. For finite gaugino mass, A or μ, the constraint planes are shifted by finite amounts. If there was any allowed solid angle for $M_{1/2} = A = \mu = 0$, the new allowed region will be a truncated cone shifted from the origin. If there was no allowed solid angle for $M_{1/2} = A = \mu = 0$, turning these parameters on can allow a finite (hyper-)polyhedron. In the absence of running, that is, if the constraints

[5]Indeed one can find more general constraints by considering an arbitrary direction in the $(H_u, \tilde{L}, \tilde{\tau}_c)$ space. We are not interested here with such a general study—all we want to point out is that μ cannot be much larger than the sfermion masses. Similar considerations apply to the A terms.

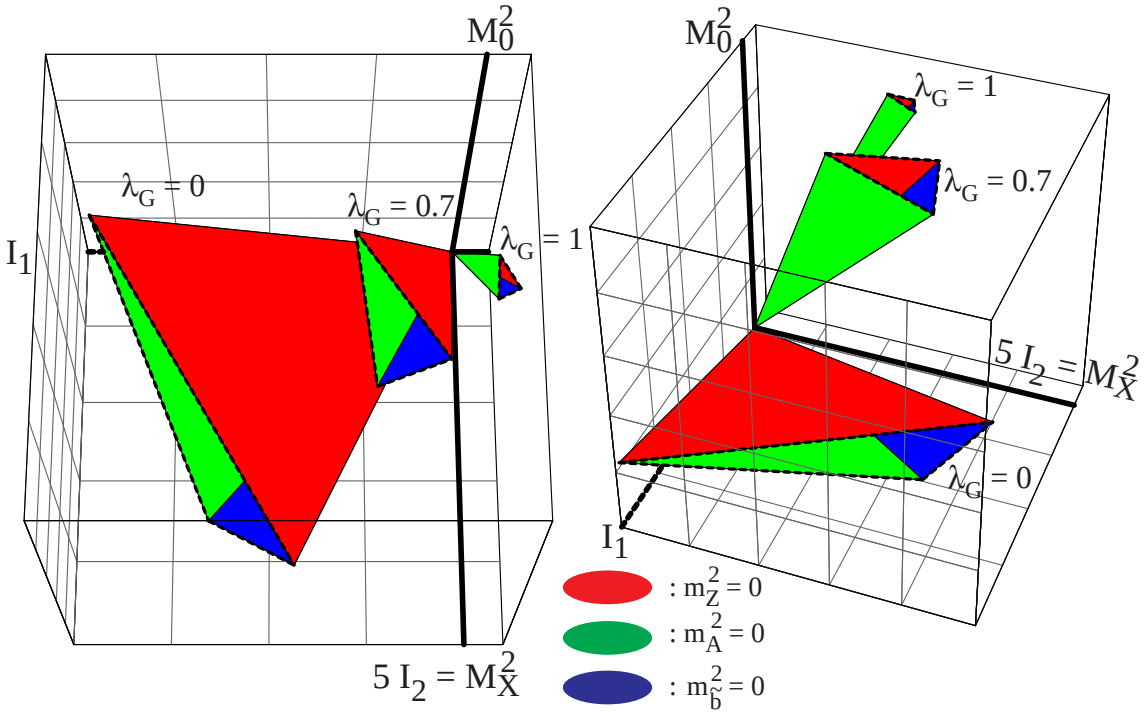

Fig. 6. The allowed "cones" in the space of scalar GUT mass parameters in the minimal SO(10) scenario with exact PQ and R symmetries, within which electroweak symmetry is correctly broken. Only the dominant (planar) constraints are shown: $m_Z^2 > 0$, $m_A^2 > 0$ and $m_{\tilde{b}}^2 > 0$. Note the focusing of the solid angle for increasing λ_G, a consequence of the exponential homogeneous evolution.

of Eqs. (25) are evaluated at the GUT scale, the cone (or polyhedron) they determine spans a solid angle of order unity. As the parameters in Eqs. (25) are evolved to lower energies, the planes turn about the origin and the cone changes; it may even close completely, in which case proper electroweak breaking becomes impossible. We will of course be interested in the cone evolved all the way to the electroweak scale; it is useful to remember that it is drawn in the space of parameters in the effective GUT-scale Lagrangian, and its boundaries correspond to those GUT-scale parameter values which lead to the vanishing of particular scalar masses at the electroweak scale or of m_2^2 at 10^5 GeV—in other words, it encompasses the GUT-scale parameters which would lead to proper breaking at the electroweak scale. A narrow cone means that it is difficult to find GUT-scale parameters which will lead to a low-energy world similar to ours.

6.2 The Homogeneous Evolution

To understand the evolution of the cone, for the time being *we restrict our attention to the homogeneous parts of the differential RG equations for the soft parameters*, neglecting the gaugino, A and μ contributions. This not only illuminates the functional behavior of the solutions, but is also directly relevant for the case we have so far espoused, the PQ- and R-symmetric one in which the μ and A parameters and the gauginos are much lighter than all the other masses (except m_Z). In this maximally-symmetric case the evolution is driven by the Yukawa couplings, which in turn depend on λ_G (and ρ_G):

- For small λ_G there is little evolution, and the cone remains wide. Here there are no generic difficulties in satisfying the constraints for a wide range of initial parameters. Whether such a range is available in particular GUT models is a question that will be answered in the next section.

- To understand the behavior for large λ_G, it is useful to change basis, considering certain fixed linear combinations $\mathcal{M}\vec{m}_i^2$ (where $\mathcal{M}$ is a constant matrix) of the seven soft-breaking parameters. Appendix A contains the solutions, in this new basis, of the RG equations for the seven soft scalar masses, as well as for the three A terms and the B parameter, in terms of integrals over only the gauge and Yukawa couplings. The matrix $\mathcal{M}$ is chosen to separate the homogeneous part of the seven scalar mass RG equations into two classes: three of the linear combinations, denoted X_t, X_b and X_τ, evolve essentially multiplicatively, contracting exponentially as they evolve down to the electroweak scale. The other four, $I_{1,2,3,4}$, are approximately invariant. (Actually one linear combination of the I_i evolves slightly due to the hypercharge D-term S discussed in Appendix A, while three other independent combinations are truly invariant; but for the present purposes we can neglect S and consider all four I_i to be invariant.) Thus we find that, for large λ_G, the three X_i are exponentially suppressed at the electroweak scale relative to their GUT-scale values, and hence, generically, also relative to the four invariants I_i. In the limit of very large λ_G, at which the Yukawa couplings tend to their "fixed points" at low energies, the constraint equations constitute seven constraints in only four variables I_i:

$$0 < \vec{m}_i^2 = \mathcal{M}^{-1} \begin{pmatrix} \vec{X} \\ \vec{I} \end{pmatrix} \sim \mathcal{M}^{-1} \begin{pmatrix} e^{-\frac{1}{\pi^2}\int \lambda^2 dt} \vec{X}_G \\ \vec{I} \end{pmatrix} \sim \mathcal{M}^{-1} \begin{pmatrix} \vec{0} \\ \vec{I} \end{pmatrix} . \tag{28}$$

  Given the value of $\mathcal{M}$ in Appendix A, it is simple to check that, in the limit $\lambda \to \infty$, there is only a trivial solution: $I_i \equiv 0$. This solution is highly nontrivial, however, in the sense that it requires adjusting the soft masses at the GUT scale such that they just cancel when combined into the I_i. At finite λ the degree of cancellation needed is just the degree to which the X_i are suppressed; schematically, then, we must adjust the GUT-scale parameters to satisfy

$$|I_i| \lesssim e^{-\frac{1}{\pi^2}\int \lambda^2 dt} X_G \,, \quad i = 1, 2, 3, 4 \,. \tag{29}$$

How much tuning is *actually* required in order to satisfy Eqs. (25)? We will determine numerically the allowed region of parameter space under various GUT-scale assumptions. Generically, one would expect a factor of $e^{-\frac{1}{\pi^2}\int \lambda^2 dt}$ for each invariant which needs a cancellation, hence $\left(e^{-\frac{1}{\pi^2}\int \lambda^2 dt}\right)^4$. However, since the leptonic Yukawa coupling is considerably smaller than the other two over much of the RG evolution, and since one of the invariants and one of the X's are essentially determined by leptonic masses, a better estimate is $\left(e^{-\frac{1}{\pi^2}\int \lambda^2 dt}\right)^3$. To alleviate this tuning problem, we should stay away from the fixed-point regime of large Yukawa couplings, or relax the PQ or R symmetries (but then $\epsilon_B \epsilon_Z$ is made worse). Away from the fixed point, a splitting $\lambda_t^G > \lambda_b^G$ also helps. In the end, of course, the question itself only makes sense in a particular context: so far we have assumed there is no theoretical bias towards any relationship between the initial parameters, except for the approximate PQ and R symmetries. In a GUT context there will be some definite biases.

The large-λ_G fine tuning for the symmetric case is illustrated graphically in Fig. 6, assuming the SO(10) scenario discussed below in which only one contracted direction X and two invariants $I_{1,2}$ play a role. The RG equations for the seven soft masses are solved for three cases: $\lambda_G = 0$ (that is, without running), an intermediate value of λ_G, and a large λ_G. The figure shows the cone in which the electroweak-breaking constraints of Eqs. (25) are satisfied, drawn in the space spanned by the three linear combinations of GUT parameters X_G, I_1 and I_2. Note that only a small number—three or four—of the constraints are the decisive ones, and once they are satisfied all others are as well; in this case, they are $m_Z^2 > 0$, $m_A^2 > 0$ and $m_{\tilde{b}}^2 > 0$. We see that near the fixed point, the cone closes up[6] around the ray $I_1 = I_2 = 0$, meaning that if we don't tune the GUT parameters to lie in this ray, then the soft-breaking parameters at the electroweak scale will not satisfy the constraints. Then either electroweak symmetry will not break ($m_Z^2 < 0$), or the potential will not be bounded from below ($m_A^2 < 0$, leading to dimensional transmutation at a scale much larger than the SUSY scale, and also to $\tan\beta \simeq 1$), or an electromagnetically-charged scalar will acquire a VEV.

The contractions may also be seen analytically in terms of sum rules, which are particular linear combinations of the electroweak-scale soft-breaking masses having only positive coefficients and chosen to be independent of all the invariants. Linear combinations of the X_i with positive coefficients can give such sum rules, for example $X_t + X_b$. These sum rules have the property that on one hand they are phenomenologically constrained to be positive (since $\vec{m}_i^2 > 0$), but on the other hand they are driven to zero by the RG equations as λ_G increases. Schematically, we have for our example

$$0 < 2m_{\widetilde{Q}}^2 + m_{\tilde{t}}^2 + m_{\tilde{b}}^2 + m_A^2 = X_t + X_b \sim e^{-\frac{1}{\pi^2}\int \lambda^2 dt}\,(X_t + X_b)_G\,. \tag{30}$$

Since each of the mass terms in this equation should be positive, for large λ_G they must each be made to evolve towards zero at low energies. Now, each mass term can itself be expressed

[6]Actually, as evident from the figure, the cones close up before they reach the ray $I_1 = I_2 = 0$. This "premature focusing" is a property of the specific GUT boundary conditions and will be discussed below.

as a sum of the X_i and the I_i; since the X_i are exponentially reduced while the I_i remain invariant, the soft masses can only evolve towards zero if the various I_i are tuned to be small already at the GUT scale. In fact, the sum rules embody the same information as the planes in the constraint equations (25) and in Fig. 6. In particular, there is a dominant sum rule corresponding to the innermost set of planes which define the constraining cone; for the case of Fig. 6, this dominant sum rule is $m_Z^2 + \frac{4}{3}(m_A^2 + m_{\tilde{b}}^2)$. It indicates which are the masses closest to saturating the constraints, and hence which are typically the lightest. Notice also that all the soft scalar masses except the sleptons appear in the sum rule of Eq. (30), indicating that essentially all these parameters contract for large λ_G. The slepton masses also contract, according to another sum rule, but to a lesser degree.

A very useful graphical way to describe the allowed domain in the space of initial parameters is to project the constraint planes onto the (hyper-)plane spanned by the various invariants. For the SO(10) scenario illustrated in Fig. 6 and described in detail below, the result is a set of lines in the plane of $(I_1/X_G, I_2/X_G)$. To normalize the axes, we define a "typical soft scalar mass" $M_0^2 \equiv \frac{1}{3}X_G = \frac{1}{3}M_{\mathbf{10}_H}^2 + \frac{2}{3}M_{\mathbf{16}_3}^2$, and use the more direct Lagrangian parameter M_X^2 (see below) instead of I_2. The horizontal and vertical axes are thus shown in units of M_X^2/M_0^2 and I_1/M_0^2, respectively. Each line forms the boundary of the half=plane where one of the m_i^2 is positive. Fig. 7 shows all eight lines, and emphasizes the region allowed by the constraint equations (24) and (25), for three values of λ_G, and for various values of λ_t^G/λ_b^G. In general the allowed region is a polygon; for the PQ- and R-symmetric SO(10) case, it is usually a triangle bounded by the lines corresponding to the three masses which appear in the dominant sum rule $m_Z^2 + \frac{4}{3}(m_A^2 + m_{\tilde{b}}^2)$. Near an edge of the triangle, the corresponding mass parameter is much smaller than all the others. The hatched region in Fig. 7(a) is where m_Z^2 is $\sim \tan\beta \sim 50$ times less than m_A^2, which is the favored scenario (as discussed in more detail below). Similar regions are indicated for the other triangles.

6.3 Evolution and Natural Selection

From the previous subsection, we conclude that if μ and $m_{\widetilde{W}}$ are chosen much smaller than the typical soft scalar masses so PQ and R are approximately valid, then we expect the allowed triangular area in the space of GUT-scale scalar masses to be small if $\lambda_G \gtrsim 0.7$ and $\lambda_t^G = \lambda_b^G$ (the focused case), and large otherwise. In fact, as we show below, both SO(10)- and SU(5)-type boundary conditions on the scalar masses result in *premature* focusing: the triangles close up for finite λ_G values, not far above unity. Within the allowed triangle, therefore, a few particles—namely those which bound the triangle itself—are very much lighter than the rest. (In contrast, for $\lambda_G \to \infty$ focusing all masses must become very light.)

Now, if there *is not* much focusing, then all the scalar masses are comparable throughout most of the triangle, while in a narrow ($\sim 1/\tan\beta$) strip within that triangle m_Z^2 is $\sim 1/\tan\beta$ times lighter than m_A^2. Tuning the GUT scalar parameters to lie within this strip suppresses the large $b \to s\gamma$ amplitudes, and allows us to generate large $\tan\beta$ with no further tuning of B and without violating LEP bounds by taking $\mu^2 \sim M_{1/2}^2 \sim m_A^2/\tan\beta$. If there *is* significant focusing, then within the small allowed triangle the particles which bound the triangle have

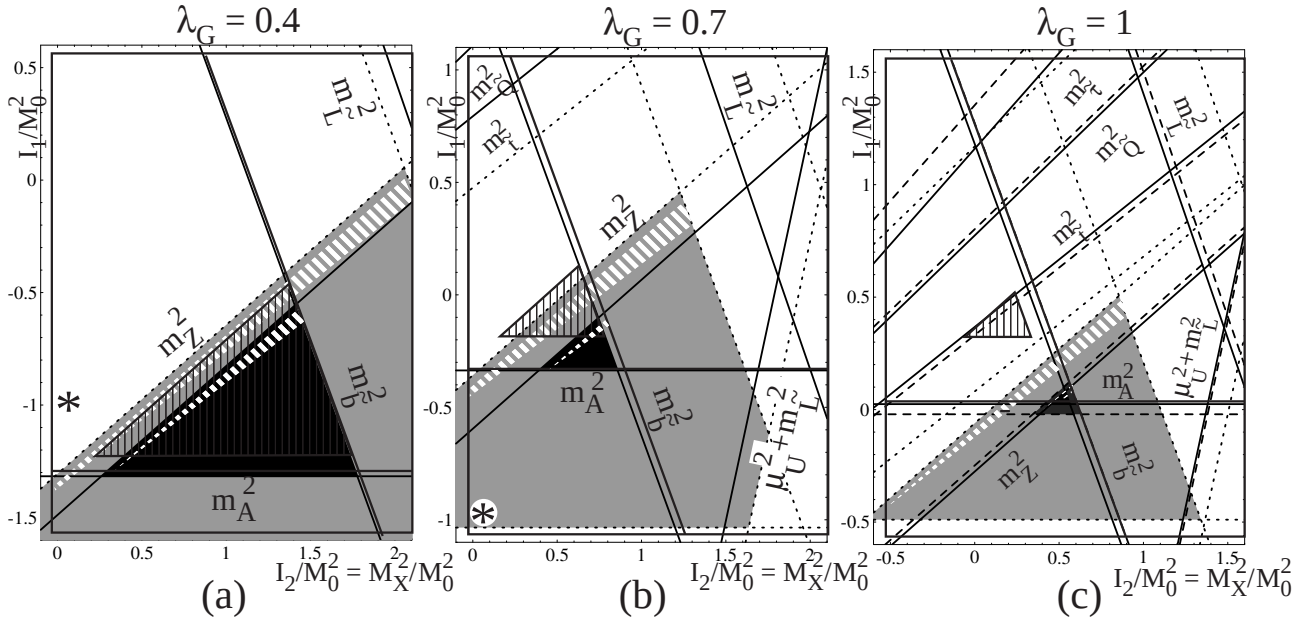

Fig. 7. The allowed regions for the same scenario as Fig. 6, but projected into two dimensions by dividing by the typical GUT squark mass M_0. The GUT-scale Yukawa coupling is increased from left to right, and both exact Yukawa unification (black areas, solid lines) and approximate unification (light gray areas, dotted lines) are considered. For $\lambda_G = 1$ we also show the perturbed case $\lambda_b^G/\lambda_t^G = 1.1$ as the dark gray area and dashed lines. The lines delineate the half plane in which the corresponding mass-squared is positive and hence acceptable. In the diagonally-hatched regions $m_Z^2 \ll m_A^2$, as discussed in the text. Finally, the vertically-hatched triangles are the allowed areas assuming exact unification and a right-handed neutrino at $M_N = 10^{12}$ GeV.

masses—indicated collectively with $m_{0,L}^2$—which are much lighter than the others, so a hierarchy $m_{0,L}^2 \ll m_{0,H}^2$ is invariably present. As we will see in Tables 3 and 4, m_Z^2 and m_A^2 are always among the light masses, while $m_{\tilde{Q}}^2$ is always one of the heavy masses. Hence throughout the triangle the dangerous $b \to s\gamma$ amplitudes are somewhat suppressed, depending on the degree of focusing. There is now no naturalness criterion to distinguish between the following situations: either the GUT scalar parameters are tuned to lie within a narrow strip in this triangle, resulting in $m_Z^2 \sim m_A^2/\tan\beta \sim m_{0,L}^2$ and requiring no further tuning of B since we can select $\mu^2 \sim M_{1/2}^2 \sim m_Z^2 \sim m_A^2/\tan\beta$ and meet the LEP constraints; or the GUT scalar parameters are not adjusted to be in the strip, so $m_Z^2 \sim m_A^2 \sim m_{0,L}^2$, but B must be tuned to $\sim 1/\tan\beta$ of its natural value because LEP requires us to select $\mu^2 \sim M_{1/2}^2 \sim m_Z^2 \sim m_A^2$. To summarize: when the PQ and R symmetries hold at the GUT scale, if there is no focusing then they hold at the electroweak scale and the overall tuning need only be $\sim 1/\tan\beta$; but if there is strong focusing then they can either hold or not hold at low scales, and in either case the need for tuning is greater than $1/\tan\beta$ by the degree of focusing.

Strictly speaking, the exactly PQ- or R-symmetric scenarios are never allowed by LEP limits, so we should in principle always take μ and $M_{1/2}$ into account in evolving the cone to low energies. But this would not qualitatively change the discussion. Clearly, approximating μ or $M_{1/2}$ by zero is valid when there is no focusing and the symmetries are approximately valid at all scales. But even when there is focusing and some of the particles end up with small masses $m_{0,L}$, introducing finite μ or $M_{1/2}$ comparable to those masses will only change the allowed triangle area and the light masses by $\mathcal{O}(1)$; in fact, if μ or $M_{1/2}$ are then increased beyond $m_{0,L}$, the focusing is alleviated such that the new value of $m_{0,L}$ is again comparable to μ or $M_{1/2}$.

If the PQ or R symmetries are not approximately valid at the GUT scale, the planes $m_i^2 = 0$ which delimit the allowed volume are shifted, and the focusing is alleviated. (The exception is when μ or $M_{1/2}$ are too big in certain GUT scenarios—then the triangles once again close for finite values of λ_G.) Irrespective of focusing, the tuning is worse than $\sim 1/\tan\beta$, as demonstrated schematically in Table 2. However, we will reserve judgment on the breaking of PQ and R symmetries until after we have studied the SO(10) and SU(5) symmetry constraints in the scalar mass sector.

7 Grand Unified Soft Masses

7.1 SO(10) and SU(5) Boundary Conditions

We now address the question of what values of initial parameters arise from theories at the GUT scale. Since the idea of Yukawa unification is largely based on symmetry principles, it behooves us to consider the implications of those same symmetries for the various soft SUSY-breaking parameters. We have already employed, in our solutions to the RG equations, the assumption that gauge coupling unification is accompanied by gaugino mass unification, in order to reduce the number of independent gaugino mass parameters to one. Let us now examine the implications of SO(10) or SU(5) symmetries for the soft-breaking masses. As a special case, we comment briefly on the universal scenario. We also examine the threshold corrections due to a light right-handed neutrino.

Consider first the simplest SO(10) scenario, in which both light Higgs doublets originate from a single $\mathbf{10}_H$ multiplet, or more generally any SO(10) model in which all the GUT fields from which the light doublets arise have degenerate soft masses. When SO(10) breaks, in general both D and F terms could split the scalar masses in a single SO(10) multiplet. D-term [39] splittings are generically present because the rank of the gauge group is reduced, but F-term splittings are more model dependent and need not arise. For example, when SUSY breaking is communicated from a hidden sector only via gravitational interactions, the soft terms are very constrained [40]. This property leads, in the minimal missing-VEV models [41], to exclusively D-term splittings. Therefore when we refer to SO(10)-type boundary conditions we will only include D-term splittings, whereas more general F-term splittings will be encompassed by the discussion of SU(5) boundary conditions, or when necessary by the completely general discussion. Accordingly, in SO(10)-type models the seven soft-breaking

masses are determined at the GUT scale by only three soft-breaking parameters: the soft Higgs mass $M_{\mathbf{10}_H}$, the third-generation soft squark and slepton mass $M_{\mathbf{16}_3}$, and a soft mass parameter M_X from a D-term that is left over when the $\mathrm{U(1)_X}$ symmetry in SO(10) is spontaneously broken. Recall that the rank of SO(10) is higher by one than that of SU(5) and of $\mathrm{SU(3)} \otimes \mathrm{SU(2)} \otimes \mathrm{U(1)_Y}$, thus $\mathrm{SO(10)} \supset \mathrm{SU(5)} \otimes \mathrm{U(1)_X}$, where the generator of $\mathrm{U(1)_X}$ is proportional to $3(\mathrm{B}-\mathrm{L}) + 4\mathrm{T_{3R}}$. One common way to break this $\mathrm{U(1)_X}$ and reduce the rank to that of the standard model is to introduce a pair of 16-dimensional Higgs representations having GUT-scale masses, $\mathbf{16}_H$ and $\overline{\mathbf{16}}_H$, which acquire VEVs in their "ν_R" components, thus preserving the SU(5) symmetry. When $\mathrm{U(1)_X}$ breaks this way, its D-term acquires a VEV proportional to the difference of the soft masses of the $\mathbf{16}_H$ and $\overline{\mathbf{16}}_H$. This D-term then contributes to the soft masses of the fields which couple to $\mathrm{U(1)_X}$ in proportion to their $\mathrm{U(1)_X}$ charges. To quantify this contribution, we define a mass parameter

$$M_X^2 = M_{\mathbf{16}_H}^2 - M_{\overline{\mathbf{16}}_H}^2 \,. \tag{31}$$

(Note that this definition differs by a factor of 10 from the definition in our previous work [16]. We chose the present definition because it is closer to the fundamental parameters of the grand-unified theory, and so is on the same footing as $M_{\mathbf{10}_H}$ and $M_{\mathbf{16}_3}$.) It is the presence of M_X which allows the up- and down-type Higgs masses to be split at the GUT scale in almost any SO(10)-unified scenario, and thus greatly facilitates proper electroweak breaking.

In SU(5), of course, the two Higgs doublet superfields originate from different representations, $\mathbf{5}_H$ and $\overline{\mathbf{5}}_H$, so their soft SUSY-breaking masses are generically split. This is also the case in some SO(10) models, for example when the light doublets are mixtures of different GUT multiplets having different soft masses, or even when they both lie in the same SO(10) multiplet but the soft terms are general enough to induce F-type splittings. Note that in this second case the Yukawa couplings are indeed exactly unified. For brevity, however, we will call any such boundary conditions on scalar masses SU(5)-type. The seven soft masses in SU(5)-type models are determined by four parameters: the two Higgs masses $M_{\mathbf{5}_H}$ and $M_{\overline{\mathbf{5}}_H}$, and the soft masses for the two representations which contain the third-generation MSSM squarks and sleptons, $M_{\overline{\mathbf{5}}_3}$ and $M_{\mathbf{10}_3}$. For comparison with the SO(10) case, we will recombine these four parameters into the same three combinations which occur in SO(10) plus an extra degree of freedom $M_{\mathrm{SU(5)}}$, as follows: $M_{\mathbf{10}_H}^2 \equiv \frac{1}{2}(M_{\mathbf{5}_H}^2 + M_{\overline{\mathbf{5}}_H}^2)$, $M_{\mathbf{16}_3}^2 \equiv \frac{3}{4}M_{\mathbf{10}_3}^2 + \frac{1}{4}M_{\overline{\mathbf{5}}_3}^2$, $M_X^2 \equiv \frac{5}{4}(M_{\overline{\mathbf{5}}_H}^2 - M_{\mathbf{5}_H}^2) + \frac{5}{4}(M_{\mathbf{10}_3}^2 - M_{\overline{\mathbf{5}}_3}^2)$, and $M_{\mathrm{SU(5)}}^2 \equiv \frac{1}{2}(M_{\mathbf{5}_H}^2 - M_{\overline{\mathbf{5}}_H}^2) + \frac{1}{2}(M_{\mathbf{10}_3}^2 - M_{\overline{\mathbf{5}}_3}^2)$. (To reiterate: in an SU(5) context, the four quantities $M_{\mathbf{10}_H}$, $M_{\mathbf{16}_3}$, M_X and $M_{\mathrm{SU(5)}}$ should be regarded just as convenient linear combinations of the underlying soft masses.)

With these redefinitions we may write the seven soft masses at the GUT scale in either SO(10) or SU(5) as follows:

$$M_U^2 = M_{\mathbf{10}_H}^2 - \frac{2}{10}M_X^2 + \frac{1}{2}M_{\mathrm{SU(5)}}^2 \tag{32}$$

$$M_D^2 = M_{\mathbf{10}_H}^2 + \frac{2}{10}M_X^2 - \frac{1}{2}M_{\mathrm{SU(5)}}^2 \tag{33}$$

$$M_{\widetilde{Q},\tilde{t},\tilde{\tau}}^2 = M_{\mathbf{16}_3}^2 + \frac{1}{10}M_X^2 + \frac{1}{4}M_{\mathrm{SU(5)}}^2 \tag{34}$$

$$M^2_{\widetilde{b},\widetilde{L}} \;=\; M^2_{\mathbf{16}_3} - \frac{3}{10}M_X^2 - \frac{3}{4}M^2_{\mathrm{SU}(5)}\,, \tag{35}$$

(Capital letters denote the parameter values at the GUT scale.) Their linear combinations X_i and I_i then take the form:

$$X_t^G \;=\; M^2_{\mathbf{10}_H} + 2M^2_{\mathbf{16}_3} + M^2_{\mathrm{SU}(5)} \tag{36}$$

$$X_b^G \;=\; M^2_{\mathbf{10}_H} + 2M^2_{\mathbf{16}_3} - M^2_{\mathrm{SU}(5)} \tag{37}$$

$$X_\tau^G \;=\; M^2_{\mathbf{10}_H} + 2M^2_{\mathbf{16}_3} - M^2_{\mathrm{SU}(5)} \tag{38}$$

$$I_1 \;=\; 2\,M^2_{\mathbf{10}_H} - 3\,M^2_{\mathbf{16}_3} + 2\,M^2_{\mathrm{SU}(5)} \tag{39}$$

$$I_2 \;=\; \frac{1}{5}M_X^2 + \frac{1}{2}M^2_{\mathrm{SU}(5)} \tag{40}$$

$$I_3 \;=\; M^2_{\mathbf{16}_3} - \frac{3}{10}M_X^2 + \frac{5}{4}M^2_{\mathrm{SU}(5)} \tag{41}$$

$$I_4 \;=\; -\,M^2_{\mathrm{SU}(5)}\,. \tag{42}$$

A common assumption in much of the previous work on unified supersymmetric theories, and in particular in large $\tan\beta$ models [13], is a "universality" of soft SUSY-breaking scalar masses at the GUT scale. Under the universality assumption, $M_{\mathbf{10}_H} = M_{\mathbf{16}_3}$ while $M_X = M_{\mathrm{SU}(5)} = 0$. We have not made this assumption because we do not expect it to hold at the GUT scale (whether or not it is a good approximation at the Planck or string scales) and because it requires [16] tremendous fine-tuning of parameters to achieve proper electroweak symmetry-breaking with large $\tan\beta$. We will have more to say about this case in the discussion in Sec. 7.2.

A much better motivated assumption is that the third-generation right handed neutrino, that is, the electroweak-singlet superfield which couples to ν_L through a Yukawa coupling with H_U, has a Majorana mass M_N smaller than M_{GUT}. Theoretically, such a mass may arise[7] from a high-dimensional operator induced at some scale $M' > M_{\mathrm{GUT}}$ (such as the string or Planck scale) and therefore be suppressed by a power of $1/M'$. Phenomenologically, a lower M_N leads through the see-saw mechanism to a higher mass for the observed neutrinos, which may then explain various cosmological and astrophysical puzzles. In any case, as long as M_N is not too far below M_{GUT}, its effects can be approximated by threshold corrections to the Yukawa couplings and to the soft SUSY-breaking scalar masses. If the SO(10)-type boundary conditions $\lambda_t^G = \lambda_N^G \equiv \lambda_G$ are valid, we can calculate the size of the corrections:

$$\Delta\lambda_\tau = \Delta\lambda_t \;=\; -\frac{1}{2}\lambda_G^3\,\frac{\ln\left(M_{\mathrm{GUT}}/M_N\right)}{8\pi^2}\,;$$

[7] This happens, for instance, in the absence of $\mathbf{126} + \overline{\mathbf{126}}$ SO(10) Higgs multiplets and of R-odd gauge singlets. Note that in order to preserve the validity of perturbation theory it is better to avoid the **126** or bigger representations, since they contribute a large positive term to the gauge β-function [20].

$$\Delta M_U^2 = -\lambda_G^2 X_t^G \frac{\ln(M_{\rm GUT}/M_N)}{8\pi^2}; \quad (43)$$
$$\Delta M_L^2 = -\lambda_G^2 X_\tau^G \frac{\ln(M_{\rm GUT}/M_N)}{8\pi^2}.$$

The consequences will be discussed in the following section.

7.2 SO(10)-type GUT Masses

We first consider the symmetric, minimal SO(10) scenario: (PQ- and R-) symmetric in that μ, A, B and the gaugino masses are negligible relative to the various squark, slepton and pseudoscalar masses; minimal in that the two light Higgs doublets are contained in a $\mathbf{10}_H$ of SO(10), so $\lambda_t^G = \lambda_b^G$ (up to small threshold corrections); and SO(10) in that the soft-breaking masses are those that arise in a minimal SO(10) theory, hence $M_{\rm SU(5)} \equiv 0$. Three independent dimensionful GUT-scale initial parameters specify the electroweak-scale consequences, and we choose them (as noted above) to be M_0^2 ($= \frac{1}{3}M^2_{\mathbf{10}_H} + \frac{2}{3}M^2_{\mathbf{16}_3} = \frac{1}{3}X_t^G = \frac{1}{3}X_b^G = \frac{1}{3}X_\tau^G \equiv \frac{1}{3}X^G$), M_X^2 ($= 5I_2$) and $2M^2_{\mathbf{10}_H} - 3M^2_{\mathbf{16}_3} = I_1$. The third invariant is a linear combination of M_0 and the first two invariants; it is not very constrained, since it is mostly associated with X_τ, which does not contract much. The fourth and last invariant vanishes identically by the SO(10) symmetry. The evolution from the GUT scale to the electroweak scale, and the contraction of the allowed triangle towards small M_X and I_1 as λ_G increases, were illustrated in Fig. 6, or in projected form (in the space of M_X^2/M_0^2 and I_1/M_0^2) in Fig. 7. Three features are worth noting:

1. In the general case of seven independent initial soft mass parameters, we expect that the allowed region closes asymptotically as $\lambda_G \to \infty$, when all four invariants must be set to zero to allow the three X_i to contract. Graphically, this means that the eight planes (or lines) corresponding to $m_i^2(\Lambda = m_Z) = 0$ and $m_2^2(\Lambda = 10^5\,{\rm GeV}) = 0$ all cross at one hyper-ray $I_{1,2,3,4} = 0$ (namely the origin in the projected space). Instead, we see that in the symmetric minimal SO(10) case the allowed region of parameter space is bounded by the planes (or lines) $m_Z^2 = 0$, $m_A^2 = 0$ and $m_{\tilde b}^2 = 0$ which cross prematurely, closing the allowed triangle for a finite value of λ_G. The reason is that SO(10) boundary conditions are compatible with $I_{1,2,3,4} = 0$ only if *all* initial masses vanish, a trivial and uninteresting scenario. For nonvanishing M_0, the restrictive SO(10) boundary conditions can only be satisfied for sufficiently small λ_G. Just how small? Consider the sum-rule-like combination $m_Z^2 + \frac{4}{3}(m_A^2 + m_{\tilde b}^2)$, which was chosen so that the invariant part is a negative definite quantity: $(-2I_1 - 7I_2 - 14I_3 + 14I_4)/183 = -4/61M_0^2$. The contracting part $(-97X_t + 216X_b + 7X_\tau)/183$ starts out positive, with a value $126/61M_0$, and contracts monotonically to zero as λ_G increases. Hence the sum $m_Z^2 + \frac{4}{3}(m_A^2 + m_{\tilde b}^2) = (-97X_t + 216X_b + 7X_\tau)/183 - 4/61M_0^2$ must vanish for a finite λ_G. The critical value turns out to be $\lambda_G \simeq 1.2$. Therefore in this minimal symmetric SO(10) case it is important that λ_G is not only small, but in particular is well below ~ 1.2, for there to be a significant allowed region of parameter space. [If we wish to

include the small contribution to this "sum rule" of the hypercharge D-term, we can again find the proper combination of masses in which the invariant part is negative and proportional to M_0^2. The result is:

$$m_Z^2 + \frac{4}{3}(m_A^2 + m_{\tilde{b}}^2) + \left[\frac{\frac{1}{6}\alpha_G \ln(M_{\rm GUT}/m_Z)}{1 + 3\alpha_G \ln(M_{\rm GUT}/m_Z)}\right](m_Z^2 + m_A^2)\,, \tag{44}$$

but there is hardly any change in the conclusions.]

2. When the parameter space contracts, it does so around a nonzero value of M_X^2/M_0^2. This highlights the important role of the $\mathrm{U(1)_X}$ D-term mass parameter in allowing both a negative m_U^2 and a sufficiently positive m_D^2, even though the RG equations drive m_D^2 down more than m_U^2.

3. Notice that the allowed cone in the PQ- and R-symmetric SO(10) scenario of Fig. 6 is confined to the $M_0^2 > 0$ half plane. This result was used when dividing by M_0^2 to determine the allowed area in the projected space of Fig. 7. We have found that, in order to satisfy the low-energy constraints of Eqs. (25) and Eq. (24), M_0^2 must always be positive for all values of λ_G and λ_t^G/λ_b^G under consideration, and for SU(5) as well as SO(10) boundary conditions on the soft scalar masses, as long as the PQ and R symmetries hold. To meet the low-energy constraints with negative M_0^2 inevitably requires very large values of $|\mu^2/M_0^2|$ and sometimes of $|M_{1/2}^2/M_0^2|$; since these would require very delicate fine-tuning of the GUT-scale parameters, we will assume $M_0^2 > 0$ for the remainder of this paper.

From Fig. 7 we can infer some properties of the universality assumption when PQ and R are valid. The asterisk $(*)$ indicates the coordinate $(M_X^2, I_1) = (0, -M_0^2)$ corresponding to universal scalar mass boundary conditions at the GUT scale. Proper electroweak symmetry-breaking occurs only if λ_t^G and λ_b^G are widely split, and then only for intermediate values of λ_G. Furthermore, to meet LEP constraints in the approximately PQ- and R-symmetric scenario, the value of λ_G must be tuned to achieve $m_Z^2 \sim \mu m_{\widetilde{W}} \ll m_0^2$. If on particular we set $\mu^2 \sim m_{\widetilde{W}}^2 \sim m_A^2/\tan\beta$ and tune λ_G with a precision $\sim 1/\tan\beta$ to get m_Z^2 light enough, we end up with the minimally fine-tuned scenario at the lower-left corner of Table 2. Thus the PQ- and R-symmetric universal case is allowed and only minimally tuned (via λ_G) *if* the Yukawas are widely split at the GUT scale.

Fig. 7 also illustrates the effects of threshold corrections due to a right-handed neutrino with a mass $M_N \ll M_{\rm GUT}$. The vertically-hatched triangles in Fig. 7 shows the area allowed when $M_N \sim 10^{12}\,\mathrm{GeV}$. The $\sim 5\%$ correction to the Yukawa couplings lowers λ_t^G and so would reduce the allowed area, but the $\sim 40\%$ negative correction to M_U^2 is far more significant and increases the area. The result is that the area remains ~ 0.1 even at $\lambda_G \simeq 1.0$, comparable to that of the dark-shaded triangle [shown only in Fig. 7(c)] which would arise from a typical $\sim 10\%$ threshhold correction to λ_t^G/λ_b^G. Notice, however, that the neutrino effect, in addition to being well motivated, also favors a small $M_X^2 \sim 0$; in other words, the ΔM_U^2 shift can substitute for the shift produced by the $\mathrm{U(1)_X}$ D-term which was needed for proper radiative

symmetry breaking. A vanishing M_X^2 could conceivably be achieved naturally by means of a symmetry. Finally, note that a light right-handed neutrino threshold is not sufficient to allow for the PQ- and R-symmetric universal case when $\lambda_t^G = \lambda_b^G$.

In Table 3 we display some characteristics of the SO(10) scenario for various values of λ_G and of λ_t^G/λ_b^G. The heavy-dashed boxes correspond to the PQ- and R-symmetric case: the three on the left are for the minimal Higgs choice ($\lambda_t^G = \lambda_b^G$), while the three on the right allow for large Higgs mixings at the GUT scale ($\lambda_t^G = 2\lambda_b^G$). The top entry in each box gives the area of the allowed triangle using the coordinates of Fig. 7, namely M_X^2/M_0^2 and I_1/M_0^2. Also shown in some interesting cases are the larger areas that would result from a slight Yukawa splitting ($\lambda_t^G = 1.1\lambda_b^G$) due to some slight mixing or threshold effect. Note that the area decreases rapidly as λ_G increases, indicating the aggravated need for fine-tuning of the GUT-scale parameters. Of course, the value of the area depends on the choice of coordinates and the metric, which are to some extent a matter of taste. We use these particular coordinates because we expect them to be a priori of order unity, and so if the triangle area is much smaller than 1 then some tuning is apparently necessary. A crossed-out box indicates that the corresponding parameter choice leads to a value for δ_b incompatible with bottom-tau unification and the low-energy values for m_b and m_τ (though some of those boxes are nevertheless filled in for reference).

The full superspectrum is completely determined, up to an overall scale, by choosing a specific point within the allowed triangle. For the PQ- and R-symmetric scenarios we are now considering, the middle of each box indicates the superspectrum that is typical near the light-m_Z^2 portion of the triangle, shown as the hatched region in Fig. 7. [Recall that, while this region is ~ 50 times smaller than the triangle and hence by definition requires that much fine-tuning, it leads to a hierarchical spectrum which is both phenomenologically allowed and requires no further tuning to achieve acceptable $\tan\beta$ and $\Gamma(b \to s\gamma)$. Furthermore, far away from this region, m_Z is greater than μ and $m_{\widetilde{W}}$ and hence is in conflict with LEP.] In most cases all the superpartners and the pseudoscalar Higgs have similar masses. However, when λ_G is large and $\lambda_t^G \simeq \lambda_b^G$, the allowed triangle is small and therefore its bounding particles, the pseudoscalar Higgs and the SU(2)-singlet bottom squark, are somewhat lighter than the other particles. And when λ_G is large but there is a large top-bottom Yukawa splitting, the SU(2)-singlet stop becomes relatively light. The last item in each box is the list of masses which can vanish simultaneously with m_Z: for the symmetric SO(10) case, they are always m_Z, m_A and $m_{\tilde{b}}$, as discussed above. In other words, we may choose parameters at the corner of the allowed triangle such that m_A and m_Z are much lower than all the other superpartner masses, or $m_{\tilde{b}}$ and m_Z are much lower than the others.

Fig. 8(a) shows contours of fixed allowed area as functions of the size of the GUT-scale Yukawa coupling λ_G and the amount of top-bottom splitting splitting λ_t^G/λ_b^G. The sharp bends occur when the $m_{\tilde{t}}^2 > 0$ constraint becomes more restrictive that the $m_{\tilde{b}}^2 > 0$ constraint, so the rate at which the triangle closes is determined by the evolution of m_Z^2, m_A^2 and $m_{\tilde{t}}^2$ rather than of m_Z^2, m_A^2 and $m_{\tilde{b}}^2$. Note the dramatic decrease in area as the maximal value of λ_G is reached for fixed λ_t^G/λ_b^G—this is the premature focusing implied by SO(10) [and SU(5), as we shall see] boundary conditions on the scalar masses. The large-λ_G, small-splitting region

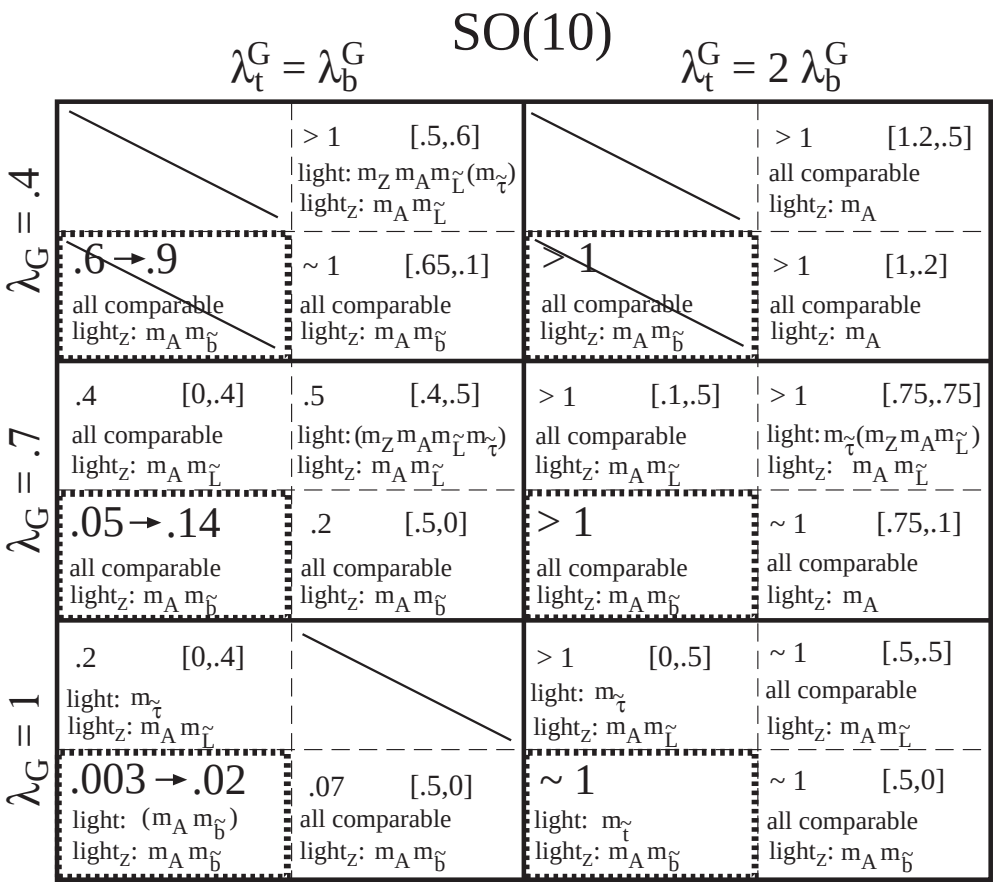

Table 3. The characteristics of exact and approximate Yukawa unification with SO(10)-type boundary conditions on the scalar masses. For each of the six choices of λ_t^G/λ_b^G and λ_G, four boxes are shown, corresponding to the presence of absence of approximate PQ and R symmetries (with μ/M_0 and $M_{1/2}/M_0$ shown respectively in square brackets). The first entry in each box is the allowed area, the second is a typical spectrum, and the third lists the masses than can be decreased simultaneously with m_Z. See the text for further details.

clearly requires very precise adjustment of the GUT-scale parameters.

We next relax either the PQ or the R symmetry, or both, and once again ask for the allowed region in the parameter space of soft scalar masses. For fixed μ and $M_{1/2}$, the planes which delimit the allowed region are now shifted by fixed amounts $\sim \mu^2$ and $\sim M_{1/2}^2$, so that they no longer intersect at the origin. We are actually most interested in the relative quantities μ/M_0 and $M_{1/2}/M_0$, that is, the amount of PQ and R breaking relative to the other (soft scalar) mass parameters. So for fixed μ and $M_{1/2}$, we should consider various slices of constant M_0 in the scalar mass parameter space. (Such slices are in fact the projections shown in Fig. 7.) Small PQ and R breaking corresponds to looking at large-M_0 slices: at such large distances from the origin, the small displacements of the planes are insignificant,

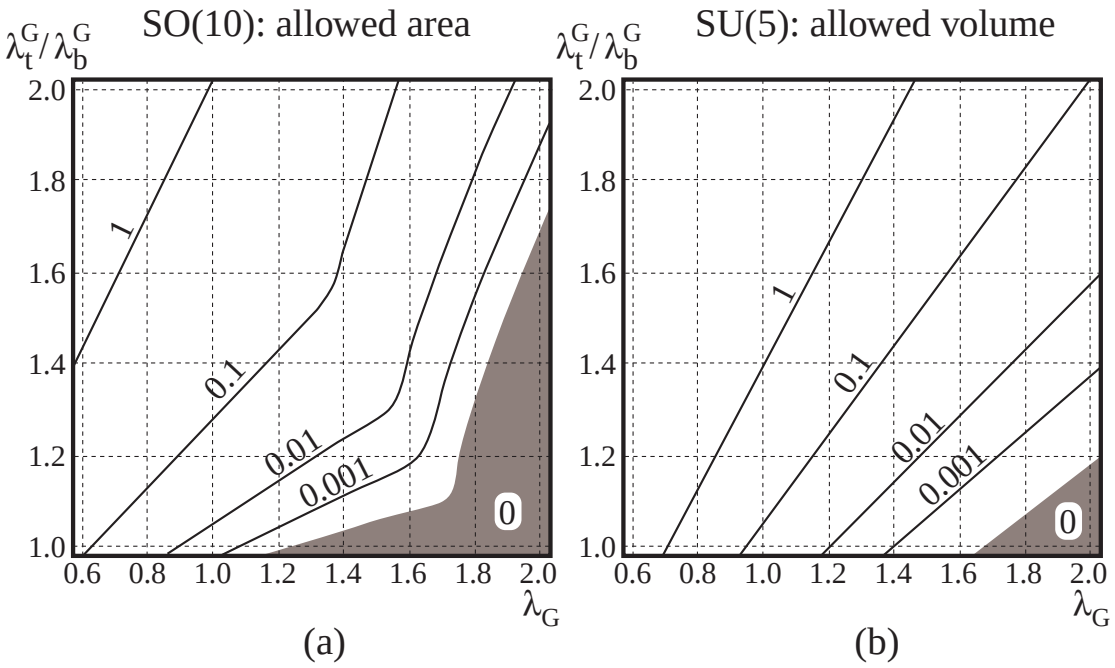

Fig. 8. Contours of constant allowed areas or volumes for SO(10)- or SU(5)-type boundary conditions, respectively, assuming exact PQ and R symmetries. Notice that small λ_G and large λ_t^G/λ_b^G values are favored, since they alleviate the premature focusing of the homogenous RG evolution.

and the allowed portion of the slice is essentially the same as in the symmetric case. Larger PQ and R breaking correspond to slices of smaller M_0, nearer the origin: the allowed portion of the slice may be large relative to the distance from the origin, in which case there is no fine-tuning, or the slice may not even intersect the allowed region, in which case those values of μ/M_0 and $M_{1/2}/M_0$ are not allowed.

We must also ensure that, within the allowed region, δ_b is within the range allowed by bottom-tau unification. [To calculate δ_b for a particular choice of μ/M_0 and $M_{1/2}/M_0$ and a given value of $\alpha_s(m_Z)$, we first determine the range of values of $m_{\tilde{b}}^2 + m_{\widetilde{Q}}^2$ within the allowed region of parameter space, and then use Eq. (5); if any resulting δ_b is acceptable, we allow that choice of μ/M_0 and $M_{1/2}/M_0$.] In Fig. 9 we show the values of μ/M_0 and $M_{1/2}/M_0$ which lead to proper electroweak symmetry breaking (i.e. there is a finite allowed triangle) and acceptable bottom-tau unification, for three representative values of λ_G, two values of $\alpha_s(m_Z)$, and either $\lambda_t^G = \lambda_b^G$ (shaded) or $\lambda_t^G = 2\lambda_b^G$ (hatched). Recall that $M_{1/2}$ is the gaugino mass at the GUT scale (which happens to roughly equal the wino mass: $m_{\widetilde{W}} = g_2^2/g_G^2 M_{1/2} \simeq 0.85 M_{1/2}$) and M_0 is the typical scalar mass also at the GUT scale, while μ is evaluated at the electroweak scale. The allowed regions are all roughly "L"-shaped. At their top and on their far right (when applicable), they are cut off by the requirement of proper electroweak symmetry breaking, while on their lower-left and upper-right sides they are bounded by the limits on δ_b.

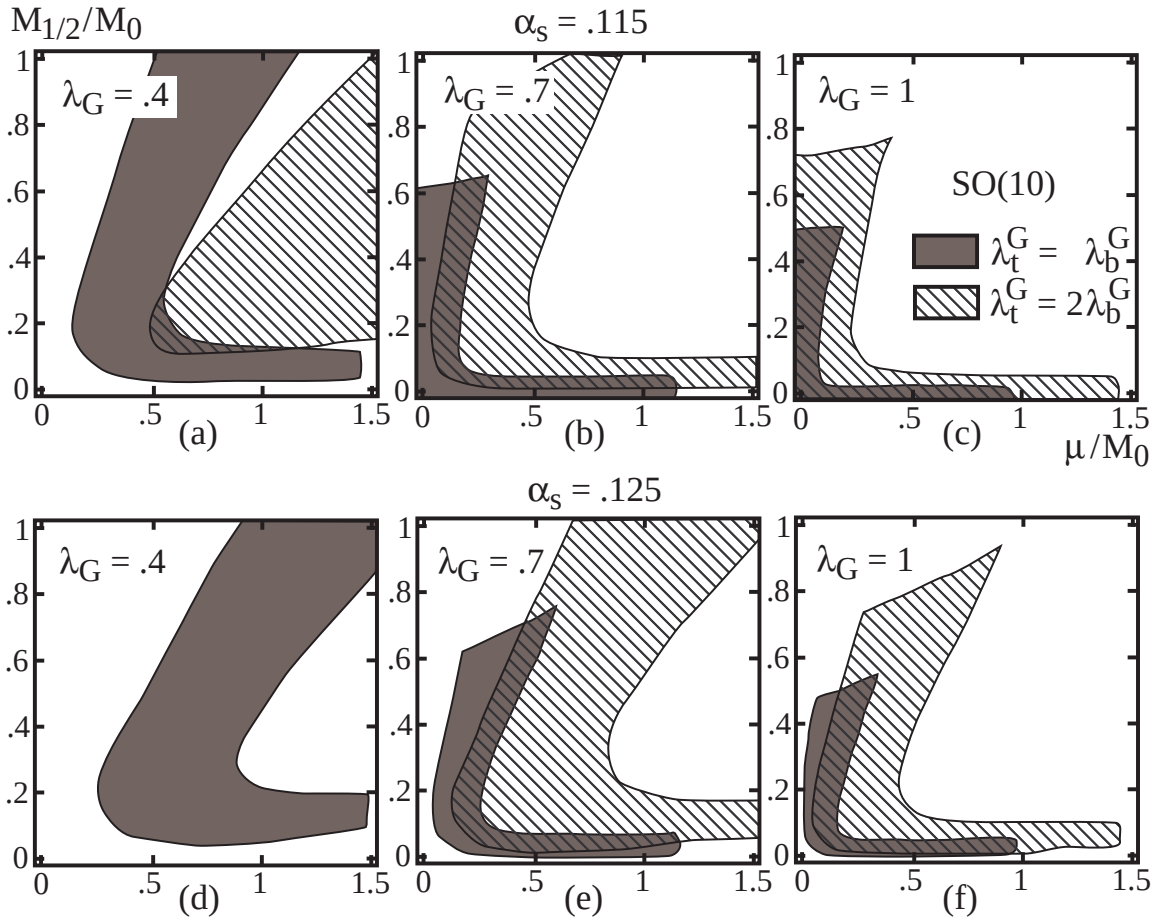

Fig. 9. The regions of GUT-scale PQ and R breaking allowed by the constraints of proper electroweak breaking and bottom-tau unification. The gray areas give the ranges of μ/M_0 (where μ is actually evaluated at a scale m_Z) and $M_{1/2}/M_0$ assuming exact Yukawa unification, while the hatched regions assume $\lambda_b^G/\lambda_t^G = 2$. All assume SO(10)-type scalar mass boundary conditions.

Particular values of μ/M_0 and $M_{1/2}/M_0$ are examined in more detail in Table 3. Once again we consider three different values of λ_G and either $\lambda_t^G = \lambda_b^G$ or $\lambda_t^G = 2\lambda_b^G$, which lead to the six major boxes of the table. Each is divided into four sub-boxes: the lower-left one is the symmetric case described above; in the lower-right one we break PQ, in the upper-left one we break R, and in the upper-right sub-box we break both the PQ and the R symmetries. (Note that these correspond spatially to the four corners of Fig. 9, as well as to Table 2.) For each sub-box we have chosen an appropriate pair of $[\mu/M_0, M_{1/2}/M_0]$ values from the allowed region of Fig. 9, and have indicated the allowed area of the triangle for those values. When choosing these values we avoided the boundaries of the allowed regions, because there the area is typically very small and hence (in some sense) unlikely. Clearly these choices are somewhat arbitrary, and there can correspondingly be some variation in the spectrum. Notice that when

the allowed area is small in the symmetric case, raising μ or $M_{1/2}$ has the expected effect of increasing the area (but recall the price one pays in $\epsilon_B \epsilon_Z$), since the focusing described above is alleviated. When either the PQ or R symmetries are approximately valid, it is still necessary to focus on the light-m_Z part of the triangle (see Table 2); when both symmetries are broken, much of the triangle is allowed by LEP and no part is selected by naturalness criteria, so we have arbitrarily chosen to look at its center. The middle entry in each sub-box indicates as before the typical superspectrum for the appropriate region of the triangle. If the various squark, slepton and pseudoscalar masses for a given sub-box are all within roughly a factor of two, we characterize them as "all comparable", and otherwise we indicate which ones are significantly lighter; masses in parenthese are only marginally lighter (i.e. somewhat less than half the heavier masses). The bottom entry of each box shows once again the particles which can become light simultaneously with the Z. If only one particle is shown, the reason is that at the other vertex of the triangle the flat-direction mass m_2^2 is negative at scales between 10^5 and 10^7 GeV, and hence that vertex would lead to improper electroweak symmetry breaking.

Returning for a moment to the universal scenario, we recall that if $\lambda_t^G = \lambda_b^G$ then the universal assumption is incompatible with even approximate PQ and R symmetries. Indeed, when the Yukawa couplings are exactly unified the universal case requires a much bigger tuning [16]. The reason is that when the scalar masses are universal the only sources of custodial breaking available for splitting m_U^2 from m_D^2 are the *small* effects of hypercharge and λ_τ. Moreover, in order to obtain $m_U^2 < m_D^2$ the gauginos must be very heavy. This can be represented by the low energy relation $m_D^2 - m_U^2 = \epsilon_c M_{1/2}^2$, where ϵ_c is small positive coefficient representing the custodial breaking induced by hypercharge. Using this relation, we see that proper symmetry breaking, i.e. $m_U^2 < 0 < m_D^2$, requires tuning some parameter (λ_G, μ^2 or $M_{1/2}^2$) to a precision of ϵ_c. Moreover, since $m_A^2 = m_U^2 + m_D^2 < \epsilon_c M_{1/2}^2$, and since $\mu \sim \mathcal{O}(M_{1/2})$ to make $m_A^2 > 0$ when the gauginos are heavy, we must tune B with a precision $\epsilon_B \sim (\mu m_{\widetilde{W}}/m_A^2)(1/\tan\beta) \sim \epsilon_c/\tan\beta$. Thus the overall tuning is at least $\sim \epsilon_c^2/\tan\beta \sim 1/\tan^2\beta$, using the rough numerical approximation $\epsilon_c^2 \sim 1/\tan\beta$. Some more tuning is required to achieve an acceptably small rate for $b \to s\gamma$. And finally, with the large δ_b corrections that result from such a spectrum, the top mass is quite light and is therefore in conflict with the recent data on m_t [26].

7.3 SU(5)-type GUT Masses

We can repeat the above analysis for SU(5)-type boundary conditions, that is, when the soft scalar masses need only be SU(5)-symmetric at the GUT scale. There are now four rather than three independent initial masses, and we will choose them to be M_0^2 [$= \frac{1}{3}M_{\mathbf{10}_H}^2 + \frac{2}{3}M_{\mathbf{16}_3}^2 = \frac{1}{6}(M_{\mathbf{5}_H}^2 + M_{\overline{\mathbf{5}}_H}^2 + M_{\overline{\mathbf{5}}_3}^2) + \frac{1}{2}M_{\mathbf{10}_3}^2$], M_X^2 [$\equiv \frac{5}{4}(M_{\overline{\mathbf{5}}_H}^2 - M_{\mathbf{5}_H}^2) + \frac{5}{4}(M_{\mathbf{10}_3}^2 - M_{\overline{\mathbf{5}}_3}^2)$], $2M_{\mathbf{10}_H}^2 - 3M_{\mathbf{16}_3}^2$ [$= M_{\mathbf{5}_H}^2 + M_{\overline{\mathbf{5}}_H}^2 - \frac{9}{4}M_{\mathbf{10}_3}^2 - \frac{3}{4}M_{\overline{\mathbf{5}}_3}^2$], and $M_{\mathrm{SU}(5)}^2$ [$\equiv \frac{1}{2}(M_{\mathbf{5}_H}^2 - M_{\overline{\mathbf{5}}_H}^2) + \frac{1}{2}(M_{\mathbf{10}_3}^2 - M_{\overline{\mathbf{5}}_3}^2)$]. The allowed region in the 3-dimensional projected space of initial scalar mass parameters is now a volume bounded by planes, which in many cases is a tetrahedron [corresponding to the allowed triangle in SO(10)]. When the PQ and R symmetries hold, we find as for SO(10) boundary conditions that the asymptotically focused case $\lambda_G \to \infty$ cannot be reached, since

the four equations $I_{1,2,3,4} = 0$ have no nontrivial solutions. Hence the allowed tetrahedron closes for a finite $\lambda_G \simeq 1.7$, which can be seen by considering the evolution of the combination $m_Z^2 + \frac{53}{34}(m_A^2 + m_{\tilde{b}}^2) + \frac{23}{34}m_{\tilde{\tau}}^2 = (-14X_t + 39X_b + 15X_\tau)/34 - 6/34M_0^2$ (neglecting the small hypercharge D-term).

Table 4 summarizes the consequences of imposing SU(5)-type boundary conditions on the soft mass parameters for various choices of λ_G and λ_t^G/λ_b^G, and for unbroken or broken PQ and R symmetries, in analogy with Table 3. This time it is the allowed volume, rather than the area, which is shown on the first line of each sub-box; also shown are the sampled values of $[\mu/M_0, M_{1/2}/M_0]$. This volume is also plotted, as a function of λ_G and λ_t^G/λ_b^G, in Fig. 8(b). The choice of coordinates in the 3-dimensional initial parameter space was as usual a matter of taste, so there is no objective way of comparing the allowed volumes in SU(5) with the allowed areas of SO(10). Qualitatively, however, it seems clear from Table 4 and Fig. 8 that SU(5)-type boundary conditions require less tuning, mainly because the additional degree of freedom $M_{\mathrm{SU}(5)}^2$ makes the Higgs splitting independent of the squark masses and hence more easily allows $m_U^2 < m_D^2$ without lowering $m_{\tilde{b},\widetilde{L}}^2$. The middle entry in each sub-box describes the typical spectrum near the $m_Z^2 = 0$ face of the allowed volume, except for cases of broken PQ *and* R in which the spectrum is shown for a generic central point in the tetrahedron. The bottom entry shows which masses are allowed to vanish simultaneously with m_Z; in general there are corners of the allowed region in which two of these may vanish along with m_Z, but which pairs may do so varies from case to case. Finally, notice that in some cases, such as $\lambda_G = 1$ and $M_{1/2} = 0$, the SU(5) entry shows that only m_A and $m_{\tilde{\tau}}$ may be light along with m_Z, whereas the SO(10) entry indicates that only m_A and $m_{\tilde{b}}$ may do so. The reason is that SU(5) allows the larger value $\mu/M_0 = 1$ indicated in that entry, for which indeed the sbottom cannot be made light but the stau can; at the smaller value $\mu/M_0 = .5$ the SU(5) boundary conditions must and do allow the sbottom to be light, since they contain the SO(10) boundary conditions as a special case.

Recall that in the PQ- and R-symmetric SO(10) analysis, in which $M_{\mathrm{SU}(5)}^2 = 0$, we were interested in the dependence of the allowed area on λ_G and λ_t^G/λ_b^G, as depicted in Fig. 8(a). The analogous contour plot of the allowed SU(5) volume is shown in Fig. 8(b). One could also ask for the volume allowed when the SO(10) scalar mass boundary conditions are perturbed in the SU(5) direction to the same extent that the SO(10) Yukawa coupling boundary conditions are relaxed: namely, restrict $|M_{\mathrm{SU}(5)}^2/M_0^2| \lesssim \lambda_t^G/\lambda_b^G - 1$ (and then normalize the volume by dividing it by $\lambda_t^G/\lambda_b^G - 1$). The answer is very simple: up to a normalization factor of order unity (due to the arbitrary definition of unit area and unit volume), the contour plot of this restricted allowed volume is similar to that of the allowed area in pure SO(10). We learn that, as we saw in the particular example of a light right-handed neutrino, small scalar mass splittings and small Yukawa coupling splittings affect the allowed area to a similar extent.

Fig. 10 shows to what extent the PQ and R symmetries may be broken. As in Fig. 8, we outline the ranges of μ/M_0 and $M_{1/2}/M_0$ which lead to a nonvanishing allowed volume in the space of initial scalar masses and to acceptable bottom-tau unification, for different choices of λ_G, λ_t^G/λ_b^G and $\alpha_s(m_Z)$. The permissible ranges extend to small μ and $M_{1/2}$ even when a fairly large δ_b is required, because within the SU(5)-allowed volume one can find corners

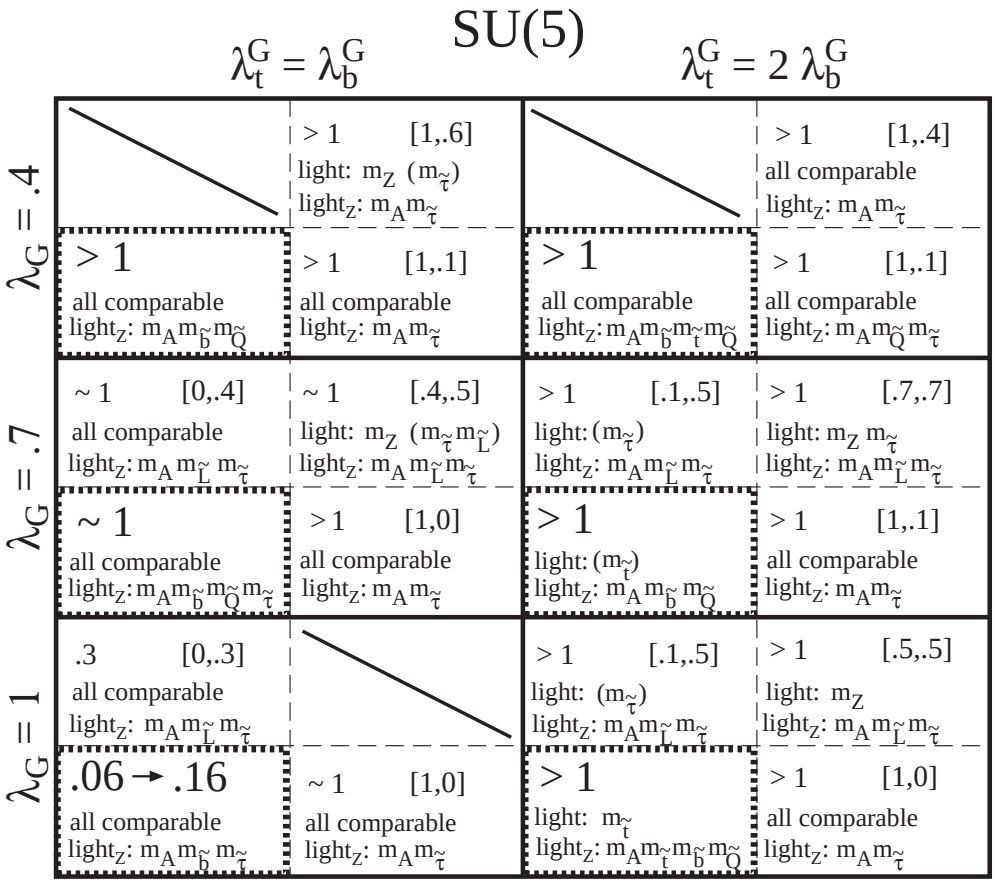

Table 4. The characteristics of Yukawa unification with SU(5)-type scalar mass boundary conditions, in analogy with Table 3.

where $m_{\tilde Q}^2 \sim m_{\tilde b}^2 \sim m_Z^2 \sim \mu^2 \sim M_{1/2}^2 \ll M_0^2$ so δ_b is quite large.

Cosmological Bounds The stability of the lightest supersymmetric particle (LSP, which we denote by χ) can lead to serious cosmological bounds on the parameters of the MSSM. In this section we discuss these bounds in the large $\tan\beta$ scenario.

As a first constraint, the LSP has to be both electrically and color neutral [42], otherwise it would have been found in searches for exotic isotopes. This is typically not a problem for us. As we have seen, in the most interesting large $\tan\beta$ scenarios, either μ or $M_{1/2}$ or both are considerably smaller than all the other SUSY parameters. In these regions of parameter space the LSP can only be a neutralino or a chargino. By a numerical study of the 4×4 neutralino mass matrix, we find that for $\tan\beta \gg 1$ and with the LEP bounds on μ and $M_{1/2}$ the LSP is always a neutralino. In the limit $|\mu| + |M_{1/2}| \gg m_Z$, this property can be easily checked by performing a perturbative diagonalization of the mass matrix.

The second constraint arises from the LSP relic mass density ρ_χ, which must not exceed the critical density of the universe today $\rho_c = (1.88 \times 10^{-29} \mathrm{g\,cm^{-3}})h^2$. We devote the rest of

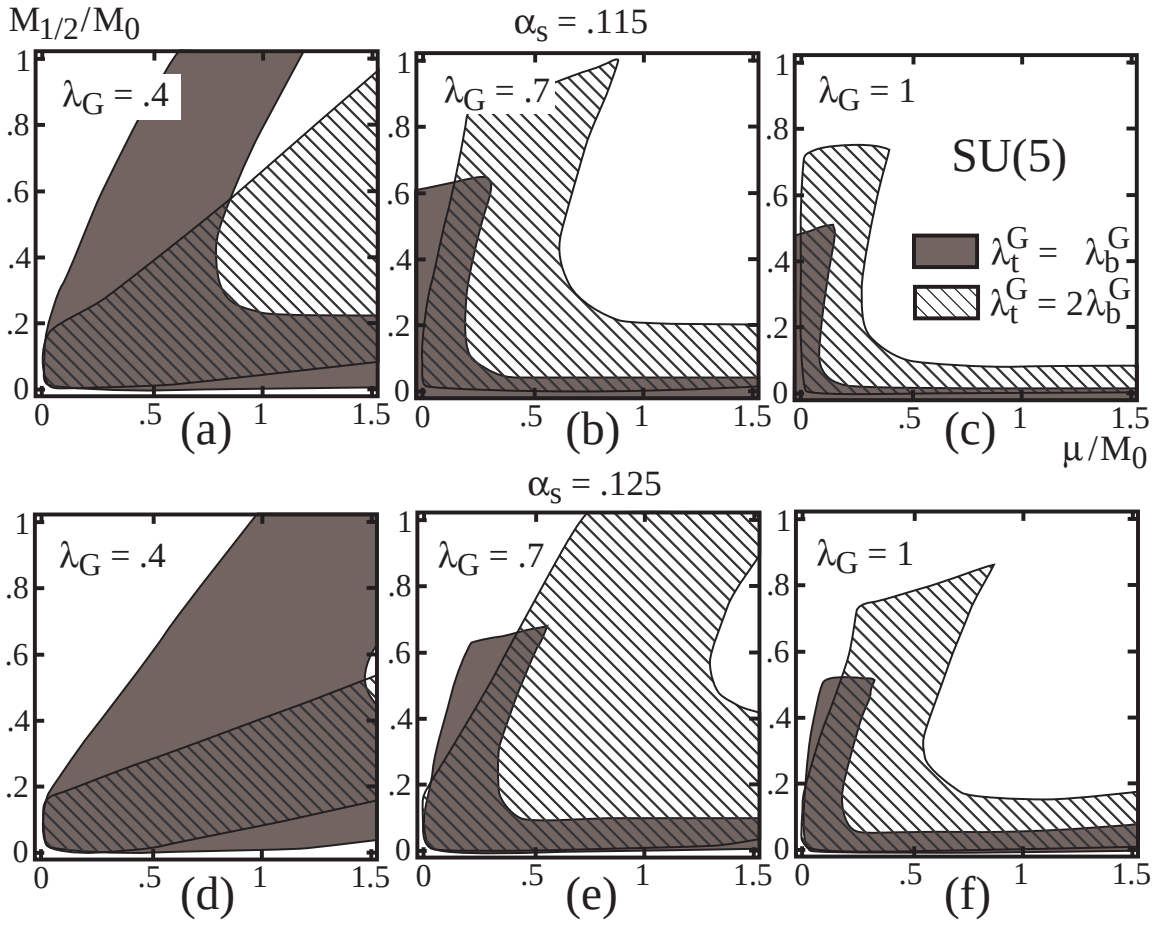

Fig. 10. The allowed regions of PQ and R breaking at the GUT scale, in analogy with Fig. 7.

this section to this issue. We base our discussion on Ref. [43], where the lightest neutralino relic density was studied but without an emphasis on the parameter regions discussed in our paper. Recently the LSP abundance in the large $\tan\beta$ scenario was also partially discussed in Ref. [32], and where our analyses overlap there is qualitative agreement.

The contribution of χ to the present Ωh^2 (where $\Omega \equiv \rho/\rho_c$) is determined by how fast the LSP annihilates when it is non-relativistic. In practice, $\Omega_\chi h^2$ is roughly inversely proportional to the annihilation cross section $\sigma_{\chi\chi}$ at a freeze-out temperature $T_F \sim m_\chi/20$ [44]. In our case the composition of χ and its annihilation properties depend crucially on μ and $M_{1/2}$. Thus the PQ and R symmetries provide once again the right language for classifying the different situations.

Let us first consider the approximately PQ-symmetric and R-asymmetric scenario $m_Z \sim \mu \ll M_{1/2}$. In this case the LSP is predominantly a Higgsino. For $m_\chi > m_W$ the annihilation into W pairs proceeds with full gauge strength via t-channel chargino exchange. The rate is easily sufficient to ensure that $\Omega_\chi h^2 \ll 1$. For $m_\chi < m_W$, one has to rely on annihilation

into standard-model fermion pairs via s-channel vector boson exchange. Now the strength of the χ couplings to the Z plays a role, as does coannihilation both with the second-lightest neutralino χ' and with the lightest chargino χ^+ (in this case via s-channel W exchange). Define $\tilde{h}^0_\pm = \tilde{h}^0_u \pm \tilde{h}^0_d$, where $\tilde{h}^0_{u,d}$ are the neutral components of the Higgsino doublets. The mass-eigenstate fields for the LSP χ and the second-lightest neutralino χ' are given by

$$\chi = \tilde{h}_+ + \mathcal{O}\left(\frac{m_Z^2}{\mu M_{1/2}}\right)\tilde{h}_-\,, \qquad \chi' = \tilde{h}_- + \mathcal{O}\left(\frac{m_Z^2}{\mu M_{1/2}}\right)\tilde{h}_+\,. \tag{45}$$

(The plus and minus signs obviously depend on our conventions.) The isospin quantum numbers of $\tilde{h}^0_u$ and $\tilde{h}^0_d$ are such that the vertices $Z\tilde{h}_+\tilde{h}_+$ and $Z\tilde{h}_-\tilde{h}_-$ vanish, while $Z\tilde{h}_+\tilde{h}_-$ has full gauge strength. For large $M_{1/2}$, Eq. (45) implies that the $\chi\chi Z$ vertex is suppressed by $\mathcal{O}\left(m_Z^2/\mu M_{1/2}\right) \sim \mathcal{O}\left(m_Z/M_{1/2}\right)$ relative to the $\chi\chi' Z$ vertex. A similar discussion applies to the coupling $\chi\chi^+ W$, which is not suppressed in this limit. Furthermore, the splittings $m_\chi - m_{\chi'}$ and $m_\chi - m_{\chi^+}$ vanish like $m_Z^2/M_{1/2}$, so χ, χ' and χ^+ are all present just before the LSP freezes out, and coannihilation is important [45]. Thus for ultraheavy gauginos the self-annihilation rate $\sigma(\chi\chi \to f\bar{f})$ is negligible but coannihilations $\sigma(\chi\chi' \to f\bar{f})$ and $\sigma(\chi\chi^+ \to \nu\bar{e}, \ldots)$ are important, since the mass splittings are very small. The result[8] is that $\Omega_\chi h^2 \ll 1$ [43]. As $M_{1/2}$ is lowered below $\sim 400\,\mathrm{GeV}$, the $Z\chi\chi$ vertex becomes important and self-annihilation becomes dominant, leading once again to $\Omega_\chi h^2 \ll 1$. At intermediate values of the gaugino mass, $\Omega_\chi h^2$ reaches a maximum $\Omega_{\max} \sim 10\%$. So the PQ-symmetric case does not suffer from an overdensity of LSP's.

As we lower $M_{1/2}$ down to $\sim m_Z$ we recover the maximally symmetric case. Now the LSP contains sizeable portions of both $\tilde{h}_+$ and $\tilde{h}_-$, and the $Z\chi\chi$ vertex has essentially full gauge strength. As a result $\Omega_\chi h^2$ is always well below 1.

Finally, let us study the effects of raising μ to arrive at an R-symmetric and PQ-asymmetric scenario $m_Z \sim M_{1/2} \ll \mu$. Now the LSP is predominantly a bino (the hypercharge gaugino): $\chi \sim \tilde{B}$. We will collectively denote the squark and slepton masses and also m_A by the single mass parameter m_0. In the limit $\mu \sim m_0 \to \infty$ the LSP is totally decoupled and $\Omega_\chi h^2$ is extremely large: all cross sections for bino annihilation vanish at least like m_χ^2/m_0^4 or like $1/(\mu^2 \tan^2\beta)$. Clearly this poses a potential problem for this scenario. For a quantitative study of the possible annihilation rates, it is useful to integrate out the heavy fields and obtain an effective Lagrangian $\mathcal{L}_{\chi\chi}$ for the bino $\tilde{B}$ and the SU(2) gauginos $\widetilde{W}_I$, $I = 1, 2, 3$. We write $\mathcal{L}_{\chi\chi} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + \mathcal{L}_4$ in two-component notation as

$$\mathcal{L}_1 = \frac{g'^2}{2\mu}\widetilde{B}H_u^\dagger\left(1 - \frac{\not{D}\overline{\not{D}}}{\mu^2} + \ldots\right)\widetilde{B} \times \tag{46}$$

$$\left(1 - \frac{D^2}{m_A^2} + \ldots\right)\left[\frac{1}{\tan\beta}H_u + \frac{1}{m_A^2}\left(\lambda_b \overline{b} b + \lambda_\tau \overline{\tau}\tau\right)\right] + \text{h.c.} \tag{47}$$

[8] For quantitative estimates we used the formulae of Ref. [43], where only the effect of coannihilation with χ' was included. We expect that accounting also for coannihilation in the charged channel [46] will lower the final value of Ω_χ, and therefore strengthens our conclusion that there the relic LSP abundance is sufficiently small.

$$\mathcal{L}_2 = \frac{g'^2}{2\mu^2}\widetilde{B}H_u^\dagger i\not{D}\left(1-\frac{\overline{\not{D}}\not{D}}{\mu^2}+\ldots\right)\overline{\widetilde{B}}H_u \tag{48}$$

$$\mathcal{L}_3 = \frac{g'^2}{2}\sum_f \frac{Y_f^2}{m_{\tilde{f}}^2}\overline{\widetilde{B}}\overline{f}\left(1-\frac{D^2}{m_{\tilde{f}}^2}+\ldots\right)\widetilde{B}f \tag{49}$$

$$\mathcal{L}_4 = \frac{gg'}{2\mu^2}\widetilde{B}H_u^\dagger i\not{D}\left(1-\frac{\overline{\not{D}}\not{D}}{\mu^2}+\ldots\right)\overline{\widetilde{W}}H_u + \text{h.c.} \tag{50}$$

where D is the covariant derivative acting to the right, Y_f and $m_{\tilde{f}}$ are respectively the $\tilde{f}$ sfermion's hypercharge and mass, $\widetilde{W} = \widetilde{W}_I\tau_I$ and τ_I are the Pauli matrices. $\mathcal{L}_{1,2,4}$ are obtained by integrating out the Higgsinos and the H_d Higgs doublet, while $\mathcal{L}_3$ arises from integrating out the sfermions. The dots represent higher derivative terms. We have approximated $m_{H_d}^2 \simeq m_A^2$. Notice that $\mathcal{L}_4$ introduces also an $\mathcal{O}(1/\mu^2)$ overlap between the LSP and the $\widetilde{W}_3$. The overall contribution of $\mathcal{L}_4$ to the LSP annihilation amplitudes is $\mathcal{O}(1/\mu^4)$. In fact the amplitude at $\mathcal{O}(1/\mu^4)$ gets contributions also from virtual bino and wino diagrams.

Let us start from the PQ- and R-symmetric scenario and increase μ above $\sim m_Z$. Then, for intermediate values of μ, the LSP annihilation cross section is determined by $\mathcal{L}_2$ and $\mathcal{L}_4$, since all other terms above are either suppressed by a large scalar mass (m_A^2, $m_{\tilde{f}}^2$) or by $1/\tan\beta$. Focusing on $\mathcal{L}_2$ and $\mathcal{L}_4$, we notice that the leading $\mathcal{O}(1/\mu^2)$ contribution to the amplitude is only in the p-wave. This contribution is determined by the first term in $\mathcal{L}_2$, and its p-wave character is easily seen by using the equations of motion in the limit of massless fermions (It is accurate to neglect fermion masses since we suppose χ to be below the top threshold in the R-symmetric scenario.) This fact is of some importance since the LSP's freeze out in the non-relativistic regime in which p-wave cross sections suffer a further suppression $\sim T_F/m_\chi$. For $m_\chi < m_W$, the LSP's annihilate into $f\bar{f}$ pairs via Z-boson exchange, at leading order from the first term in $\mathcal{L}_2$. We find that the relic density $\Omega_\chi h^2$ exceeds 1 already for $\mu > 200 - 300\,\text{GeV}$ (depending on how close m_χ is to $\frac{1}{2}m_Z$). The situation does not improve when $m_\chi > m_{W,Z}$ and all the bosonic channels WW, ZZ, Zh and hh are open and dominant. There are two contributions to these channels: a p-wave term $\sigma_p \sim 1/\mu^4$ from the first term in $\mathcal{L}_2$, and an s-wave term $\sigma_s \sim 1/\mu^8$ from the second term in $\mathcal{L}_2$ and from virtual gaugino diagrams involving the first terms in $\mathcal{L}_2$ and $\mathcal{L}_4$. (As already stated we use $\mathcal{L}_2$ to explicitly display the suppression of s-wave processes, a fact also observed in Ref. [43]. For quantitative estimates, we calculated σ_s by returning to the partial wave amplitudes given in Ref. [43].) In the limit $m_\chi \gg m_Z$, σ_p is dominated by final states with zero helicity (longitudinal vector bosons and scalars). The amplitude in this case is readily given by the annihilation into Goldstone and Higgs bosons from the first term in $\mathcal{L}_2$ via the equivalence theorem. For the purpose of our qualitative discussion we only kept this leading term. (We expect the complete result not to be drastically different in the region $m_\chi \sim m_Z$; in fact we checked this explicitly for $\chi\chi \to WW$ by using the formulae in Ref. [43].) Using these estimates for $\sigma_{s,p}$ we find, again, that $\Omega_\chi h^2 > 1$ for $\mu \gtrsim 250\text{GeV}$. We are thus led to the interesting conclusion that we cannot essentially move away from the PQ- and R-symmetric scenario towards the PQ-asymmetric one, if all the other superpartners and the pseudoscalar Higgs are very heavy. Nonetheless the

moderate $\mu \sim 200 - 300$ GeV scenario is interesting, since in this case the LSP could account for all the dark matter and give $\Omega = 1$.

When μ exceeds $\sim 300\,\mathrm{GeV}$ some other particle (namely the pseudoscalar Higgs or some sfermion) has to be lighter in order to avoid an overdensity of LSP's. A quick inspection of the above effective Lagrangian shows that, by lowering m_A^2 or some $m_{\tilde{f}}^2$, only the annihilation into $f\bar{f}$ can be significantly affected [the amplitude into bosons from $\mathcal{L}_1$ is $\mathcal{O}(1/\tan\beta)$]. This process can be mediated by (I) t-channel $\tilde{f}$ exchange via $\mathcal{L}_3$, or (II) s-channel pseudoscalar Higgs exchange via $\mathcal{L}_1$. In case (I) the amplitude is p-wave and the effective vertex is $\sim 1/m_{\tilde{f}}^2$. Supposing a sfermion $\tilde{f}$ is relatively light, we get the following estimate for the bino-like LSP relic density:

$$\Omega_\chi h^2 \simeq \left(\frac{100\,\mathrm{GeV}}{m_\chi}\right)^2 \left(\frac{m_{\tilde{f}}}{r_f^{1/4} Y_f 100\,\mathrm{GeV}}\right)^4 \left(\frac{x_F}{25}\right)^2 \tag{51}$$

where Y_f is the fermion hypercharge (so $Y_Q = 1/3$) and r_f is the dimension of the f multiplet (so $r_Q = 6$). Thus, leaving all the other fine tunings untouched, an acceptable Ω requires an additional tuning $m_{\tilde{f}}^2$ to at least an order of magnitude below its natural scale. The amplitude for case (II) behaves like $1/\mu m_A^2$, but is in the s-wave. In this case only the final states $\bar{b}b$ and $\bar{\tau}\tau$ are relevant (the lighter fermions are suppressed by the small Yukawa couplings). We estimate:

$$\Omega_\chi h^2 \simeq \left(\frac{100\mathrm{GeV}}{m_\chi}\right)^2 \left(\frac{\mu}{350\,\mathrm{GeV}}\right)^2 \left(\frac{m_A}{350\,\mathrm{GeV}}\right)^4 \left(\frac{x_F}{25}\right) . \tag{52}$$

(We are being a little sloppy in the above equations, by neglecting m_χ^2 terms in the sfermion and pseudoscalar Higgs propagators, but the conclusions would not be changed much by a more careful computation) Notice that even though the above Ω scales like m_0^6, the result is comparable to that in Eq. (51). This is partially due to the s-wave enhancement. We see once again that if μ is increased from its minimally-allowed value, then some mass parameter (m_A^2 in this case) must be made light to meet cosmological bounds. Typically, such a requirement entails some further tuning of some GUT-scale parameters.

Which annihilation channel is more likely? Assuming SO(10) boundary conditions on the scalar masses, Table 3 shows that the only particles that can be made very light (i.e. comparable to m_Z) are the pseudoscalar Higgs and the sbottom, while for SU(5)-type boundary conditions Table 4 adds the stau to this list. Hence efficient fermion production through sfermion exchange [case (I)] requires making $m_{\tilde{b}} \sim 100\,\mathrm{GeV}$ or $m_{\tilde{\tau}} \sim 200\,\mathrm{GeV}$ (under SU(5)-type conditions). Efficient fermion production through pseudoscalar Higgs exchange [case (II)] need not require m_A to be quite as light (depending on μ), but recall that lowering m_A also increases the need to tune B in order to generate large $\tan\beta$. So both channel are roughly equally unlikely. There is, of course, the possibility that the first- or second-generation sfermions are light—their initial values need not be related to those of the third generation, and their evolution is essentially decoupled from the third generation and Higgs sectors—in which case they could remedy the difficulties with the R-symmetric, PQ-asymmetric scenario.

We conclude that whenever μ is small, whether the gauginos are light or heavy, the cos-

mological density of the LSP is well below critical. In the large μ but light gaugino case, LSP annihilation is unacceptably suppressed if all the other superpartner and pseudoscalar Higgs masses are large. The annihilation rate can be sufficient if some of those masses are lowered, either through fine-tuning of m_A, $m_{\tilde{b}}$ or $m_{\tilde{\tau}}$, or perhaps by appealing to the yet-unspecified first two generations of squarks and sleptons. Needless to say, the course nature has chosen will be definitively revealed by future measurements of the superspectrum.

8 Conclusions

In this paper we have studied some of the consequences of large third-generation Yukawa couplings in the minimal supersymmetric extension of the standard model, subject to various grand unification assumptions. We have focused our attention on the two cases $\lambda_\tau^G = \lambda_b^G = \lambda_t^G$ and $\lambda_\tau^G = \lambda_b^G \sim \lambda_t^G$ imposed at the GUT scale $\sim 10^{16}\,\mathrm{GeV}$, but most of the conclusions are qualitatively unchanged if these conditions are instead enforced at the Planck or string scales, and a few general features remain even if one only assumes that $\lambda_\tau \sim \lambda_b \sim \lambda_t$ at some very high scale. For example, the need to tune some parameters to at least one part in fifty ($\sim m_t/m_b$) is a generic consequence of LEP bounds and the structure of the MSSM Lagrangian. This is also perhaps the most bothersome conclusion: as long as the top and bottom Yukawa couplings start out comparable at the GUT scale, there is no way to both explain the top-bottom mass hierarchy naturally (in the technical sense) and avoid tuning $m_Z^2 \ll m_S^2$. But if this one bitter pill is grudgingly swallowed, the remaining features of the large $\tan\beta$ scenario are intriguing. Let us assume for the moment that there is no theoretical bias about physics at the GUT scale other than the existence of a GUT. Then, if we ask that an $\mathrm{SU(3)_c} \times \mathrm{U(1)_{em}}$-invariant vacuum as well as the experimental rate of $b \to s\gamma$ are to be typical rather than unlikely outcomes of the GUT-scale parameters, and furthermore that the prediction for m_b/m_τ agree with its experimental value as extracted using QCD sum rules, then:

(I) the Lagrangian at the GUT scale should display approximate PQ and R symmetries;

(II) the value of the unified Yukawa coupling should either be $\lambda_G \simeq 0.6 - 0.7$ if the Yukawas are exactly unified ($\lambda_b^G = \lambda_t^G \equiv \lambda_G$), or be $\lambda_G \gtrsim 0.8$ if the Yukawas are significantly split (for example $2\lambda_b^G = \lambda_t^G \equiv \lambda_G$);

(III) if the Yukawas are exactly unified, the soft SUSY-breaking scalar masses at the GUT scale should be SU(5)- but not SO(10)-type, while if the Yukawas are split then they can be of either type; and

(IV) threshold corrections to the $\mathrm{SU(3)_c}$ gauge coupling at the GUT scale must be significant and negative relative to the $\mathrm{SU(2)} \times \mathrm{U(1)_Y}$ couplings.

All of these features have phenomenological, testable consequences. We expect:

(I) light charginos and neutralinos, which may furnish tantalizing signals at LEP II and would definitely be seen at the LHC; and large masses for the squarks and sleptons (at

least of the third generation), the pseudoscalar Higgs and the charged Higgs bosons [see Eq. (22)];

(II) either a top mass between 160 and 170 GeV if the Yukawas are exactly unified, or $m_t \gtrsim 175\,\mathrm{GeV}$ if they are significantly split at the GUT scale; the Yukawa splitting at M_{GUT} is reflected in the superspectrum, for example in a light stop, or better yet in the mass combination $(X_t - X_b)/(X_t + X_b)$ defined using Appendix A, which is very sensitive to both λ_t^G/λ_b^G splitting and to the departure from SO(10) boundary conditions on the soft scalar masses (see the next point);

(III) a large value for the (almost) invariant combination $I_4 = \mu_Z^2 + \mu_A^2 + \frac{3}{2}m_{\tilde{t}}^2 - \frac{3}{2}m_{\widetilde{Q}}^2 - \frac{1}{2}m_{\tilde{\tau}}^2 + \frac{1}{2}m_{\widetilde{L}}^2$, at least if the Yukawas are exactly unified, since SO(10)-type boundary conditions have vanishing $M_{\mathrm{SU}(5)}^2 = -I_4$; and

(IV) a value of $\alpha_s(m_Z) \simeq .115$ somewhat below the central gauge unification prediction.

In particular, if feature (I) is actually borne out by future discoveries, that is, if $m_Z^2 \ll m_S^2$ (where m_S is a typical soft-breaking scalar mass) is transformed from an unnatural assumption to simply an experimental fact, then the large $\tan\beta$ scenario is as natural as the small $\tan\beta$ conventional one. The two offer very different explanations of the top-bottom mass hierarchy. But the large $\tan\beta$ scenario offers a more robust test of the bottom-tau unification hypothesis, since if λ_b and λ_τ are $\mathcal{O}(1)$ at the GUT scale they should be much less subject to perturbations by other operators. In other words, the conventional scenario suffers from uncertainties in bottom-tau unification from physics at remote scales, while with large $\tan\beta$ the uncertainties are at low energies and hence are imminently accessible. As a result, only for large $\tan\beta$ is there a tight relationship between m_t, m_b, m_τ and the superspectrum.

It is important to note that some of the above predictions strongly depend on the allowed range for the bottom mass. The uncertainty in this mass is dominated by our estimate of the theoretical error in the QCD sum rule extraction. If, for example, we would know that $m_b(m_b) < 4.15\,\mathrm{GeV}$, then the top mass would be at least 170 GeV for exact Yukawa unification or at least 180 GeV for split Yukawas, and the former would be disfavored because of the tuning mandated by its large λ_G. The upper bound on $\alpha_s(m_Z)$ would also be strengthened, and would rely less on fine-tuning arguments. Of course, the experimental uncertainty on $\alpha_s(m_Z)$ must be reduced to convincingly test these predictions. But it is conceivable that in the next decade we will know the superspectrum well enough to calculate δ_b and the logarithmic threshold corrections; if we can also extract the bottom mass to $\mathcal{O}(\alpha_s^2)$ and measure the top mass to within a GeV, then precision tests of Yukawa unification at the GUT scale would be within our reach.

After this work was essentially completed, the CDF and D$\emptyset$ collaborations at Fermilab announced the long-awaited discovery of the top quark [26]. The PQ- and R-symmetric scenario we have advocated predicts a top mass that agrees very well with the values determined by these experiments: whereas we predict $160\,\mathrm{GeV} < m_t \lesssim 190\,\mathrm{GeV}$ (using the approximate fixed-point value as an upper bound), CDF measures $m_t = 176 \pm 8 \pm 10\,\mathrm{GeV}$, while D$\emptyset$ makes

the less precise determination $m_t = 199^{+19}_{-21} \pm 22\,\mathrm{GeV}$. The argument can of course be reversed: the measured value of the top mass lends further support to a PQ- and R-symmetric Lagrangian. As the uncertainty in m_t is reduced, it will serve as an increasingly powerful test to distinguish the various scenarios we have considered.

How did we arrive at the most likely set of parameters? As sketched in Table 2, if the hierarchy $v_U \gg v_D$ and the suppression of $\Gamma(b \to s\gamma)$ are to be obtained most naturally, the PQ and R symmetries should be approximately valid but without making the typical superpartners too heavy nor the charginos and neutralinos too light: the most desirable superspectrum hierarchy is $\mu/m_0 \sim m_{\widetilde{W}}/m_0 \sim 1/7$. (In highly-focused situations, such as SO(10) with $\lambda_t^G = \lambda_b^G \gtrsim 1$, there can be two mass hierarchies, but since such situations are always more fine-tuned they are not presently relevant.) From Fig. 3 and Table 1, we learn that the resulting value of δ_b ($\sim 5\%$) is compatible with bottom-tau unification only if: (1a) $\lambda_G \gtrsim 0.6$ and $m_t \gtrsim 160\,\mathrm{GeV}$ for $\alpha_s \simeq 0.115$ or (1b) $\lambda_G \gtrsim 1$ and $m_t \gtrsim 180\,\mathrm{GeV}$ for $\alpha_s \simeq 0.125$, if the Yukawas are exactly unified; or (2) $\lambda_G \gtrsim 0.8$, $m_t \gtrsim 175\,\mathrm{GeV}$ and $\alpha_s \sim 0.115$ if the Yukawas are significantly split at the GUT scale. Then turning to Fig. 8, we conclude that case (1a) can be saved from further tuning by allowing SU(5) scalar mass boundary conditions and keeping λ_G below roughly 0.7, but case (1b) would always require large tuning because of its large λ_G. The approximately-unified case (2), on the other hand, can be naturally obtained by either SU(5)- or SO(10)-type scalar mass boundary conditions. Scenarios (1a) and (2) are therefore the two we have proposed as the most likely in the absence of more specific model-building biases.

What GUT models would yield these preferred scenarios? Our original motivation for studying unified Yukawa couplings was provided by the minimal SO(10) scenario, in which both light Higgs doublets lie in the same $\mathbf{10}_H$ multiplet. This is the case, for example, in the simplest implementation of the Dimopoulos-Wilczek missing VEV mechanism [41] for solving the doublet-triplet splitting problem. In such models, the soft SUSY-breaking parameters which remain after integrating out the heavy GUT sector can be rather constrained. This is indeed the case when SUSY breaking is communicated to the GUT sector only via gravitational interactions with a hidden sector. Consequently, the structure of the soft terms it tightly linked (see for example Ref. [40]) to the GUT superpotential. It can then be shown that the only source of Higgs splitting for minimal missing VEV models is the D-terms, so the SO(10)-type scalar mass boundary conditions hold. Therefore, to allow the freedom of SU(5)-type boundary conditions favored in scenario (1a) above while preserving the unified Yukawa relations $\lambda_t^G = \lambda_b^G = \lambda_\tau^G$, more general soft terms would be required. With such terms, as may be produced when there is moduli field dependence of the GUT superpotential couplings [47], it may be possible to induce additional, F-type splittings between the MSSM particles. If scenario (1a) were supported experimentally, it would thus shed some light on the mechanism which breaks supersymmetry. On the other hand, when the Yukawa couplings are split at the GUT scale, as in the second favored scenario above, the tuning can always be made minimal by using the PQ and R symmetries. Split Yukawas would be completely expected in SU(5) models or as a consequence of string theory, but they could also arise in SO(10) models when the light Higgs doublets originate in several SO(10) multiplets. Note that even with

universal GUT scalar masses the fine-tuning can be minimal (that is, $\sim 1/\tan\beta$) if λ_t^G/λ_b^G is significantly greater than unity.

If we do espouse some particular class of models, we may be willing to accept a scenario which requires tuning to better than one part in fifty. For example, the simplest SO(10) scenario with D-terms as the only GUT source of Higgs doublet splittings requires roughly an extra order of magnitude in tuning. But theoretically it is appealing for its simplicity, and its tuning can be somewhat mitigated by the phenomenologically-favored light right-handed neutrino (which for large $\tan\beta$ does not impair bottom-tau unification [15, 19, 20]). Moreover, if a model were sufficiently predictive to specify the GUT-scale boundary conditions in terms of few unknown parameters or even none, then the conditions we have specified for proper electroweak symmetry breaking would either be fulfilled—in which case the model would not be fine-tuned but rather remarkably predictive—or not fulfilled, in which case it would be ruled out. We do not know of any such models at present; until a candidate is found, we can only offer arguments of naturalness to point us in the right direction.

If we are willing to sacrifice some naturalness, then also the PQ or R symmetries may be relaxed. Note that relaxing both would lead to a large δ_b and hence a light top, in contradiction with the recent measurements of CDF and DØ; therefore the typical SUSY-breaking scalar mass *must* be significantly above m_Z. Also, without these symmetries other flavor-changing neutral current processes beyond $b \to s\gamma$ may be problematic. In any case, the requirement of proper electroweak breaking does not significantly favor one symmetry over the other. Cosmological upper bounds on the relic LSP density, however, favor a spectrum that is only PQ-symmetric over one that is only R-symmetric. This is of benefit to models (see for example Ref. [48]) in which the μ problem is solved by generating μ radiatively from gaugino masses or A terms, typically leading to $\mu \sim \alpha M_{1/2}$. On the other hand, with light gauginos and large μ the predominantly-bino LSP annihilates inefficiently and "overcloses" the universe. To reconcile the predicted LSP relic abundance with the measured age of the universe, one of the superpartners or the pseudoscalar Higgs must be tuned light, or else PQ symmetry must be partially restored by lowering μ down to $200 - 300\,\mathrm{GeV}$. Of course, at the edge of the allowed range for these parameters, the LSP is a prime candidate for the dark matter.

There are many aspects to the Yukawa-unified MSSM beyond the question of naturalness. In Appendix A we give the exact, semi-analytic solution to the complete 1-loop RG equations for the third generation and the Higgs sector. Understanding their behavior under various assumptions and boundary conditions was a prime topic in our study. Secs. 7.2 and 7.3 addressed the relationship between these boundary conditions and the superspectrum, and presented the ranges in which μ and $M_{1/2}$ must fall to allow proper electroweak symmetry breaking. The effects of a light (relative to M_{GUT}) right-handed neutrino threshold were examined and found helpful to symmetry breaking, while it is known that they have negligible impact on $b - \tau$ unification at large $\tan\beta$. The process $b \to s\gamma$ was reexamined, and the possibility of cancellations between various diagrams enhanced by large $\tan\beta$ were identified for the first time. Finally, various issues regarding proper electroweak breaking were raised and resolved: the two flat directions which could destabilize the scalar potential, and the scales at which they could pose a danger; the constraints on the trilinear A parameters even

for third-generation sfermions due to the hierarchy $m_Z^2 \ll m_S^2$; and similar constraints on the trilinear μ couplings which are often neglected.

The implications of a hierarchy in the Higgs expectation values rather than in the third-generation Yukawa couplings are surprisingly rich. Many aspects of the MSSM are qualitatively changed by this assumption, and the phenomenological consequences of these changes are clear and accessible to the next generation of accelerator (and perhaps dark matter) experiments. Therefore the large $\tan\beta$ scenario offers a qualitatively different alternative to the often-used small $\tan\beta$ "standard" supersymmetric model. We have used criteria of naturalness to distinguish between the various options for achieving large $\tan\beta$. Admittedly, these criteria have also revealed that all large $\tan\beta$ models appear to require some fine-tuning of the GUT-scale parameters which may not be needed for small $\tan\beta$. In other respects, however, such as bottom-tau unification, Yukawa unification has distinct advantages over the conventional paradigm. And in the near future, most questions of naturalness will be replaced by solid experimental data, which will be the final arbiter of all $\tan\beta$ scenarios, large and small.

Acknowledgments

We would like to thank L.J. Hall, R. Hempfling, C. Kolda, M. Olechowski and F. Zwirner for stimulating conversations on various topics in this study, and the Aspen Center for Physics where some of this work was done. U.S. would also like to thank the Theory Group at CERN for its hospitality during the final stages of this paper. This work was supported in part by the National Science Foundation under grants PHY-91-21039 (R.R.) and PHY-8611280 (U.S.).

Appendix A: Solving the RG Equations

The 1-loop RG equations for the parameters of the MSSM are recounted below. We use the notation $\frac{d}{d\tau} \equiv -8\pi^2 \frac{d}{d\ln\mu}$ where μ is the mass scale, as well as $m_Z^2 = -2m_U^2 = \mu_Z^2 - 2\mu^2$ and $m_A^2 = m_U^2 + m_D^2 = \mu_A^2 + 2\mu^2$ where $m_U^2 = \mu_U^2 + \mu^2$ and $m_D^2 = \mu_D^2 + \mu^2$ are the up- and down-type Higgs mass parameters in the scalar potential. The soft SUSY-breaking parameters run according to

$$\begin{aligned}
\frac{d}{d\tau}\mu_U^2 &= -3\lambda_t^2 X_t^A - \frac{1}{2}\frac{3}{5}g_1^2 S + \frac{3}{5}g_1^2 M_1^2 + 3g_2^2 M_2^2 \\
\frac{d}{d\tau}\mu_D^2 &= -3\lambda_b^2 X_b^A - \lambda_\tau^2 X_\tau^A + \frac{1}{2}\frac{3}{5}g_1^2 S + \frac{3}{5}g_1^2 M_1^2 + 3g_2^2 M_2^2 \\
\frac{d}{d\tau}m_{\tilde{t}}^2 &= -2\lambda_t^2 X_t^A + \frac{2}{3}\frac{3}{5}g_1^2 S + \frac{16}{15}g_1^2 M_1^2 + \frac{16}{3}g_3^2 M_3^2 \\
\frac{d}{d\tau}m_{\tilde{b}}^2 &= -2\lambda_b^2 X_b^A - \frac{1}{3}\frac{3}{5}g_1^2 S + \frac{4}{15}g_1^2 M_1^2 + \frac{16}{3}g_3^2 M_3^2 \\
\frac{d}{d\tau}m_{\widetilde{Q}}^2 &= -\lambda_t^2 X_t^A - \lambda_b^2 X_b^A - \frac{1}{6}\frac{3}{5}g_1^2 S + \frac{1}{15}g_1^2 M_1^2 + 3g_2^2 M_2^2 + \frac{16}{3}g_3^2 M_3^2 \\
\frac{d}{d\tau}m_{\tilde{\tau}}^2 &= -2\lambda_\tau^2 X_\tau^A - \frac{3}{5}g_1^2 S + \frac{12}{5}g_1^2 M_1^2 \\
\frac{d}{d\tau}m_{\widetilde{L}}^2 &= -\lambda_\tau^2 X_\tau^A + \frac{1}{2}\frac{3}{5}g_1^2 S + \frac{3}{5}g_1^2 M_1^2 + 3g_2^2 M_2^2
\end{aligned} \tag{53}$$

and

$$\begin{aligned}
\frac{d}{d\tau}A_t &= -6\lambda_t^2 A_t - \lambda_b^2 A_b + \frac{13}{15}g_1^2 M_1 + 3g_2^2 M_2 + \frac{16}{3}g_3^2 M_3 \\
\frac{d}{d\tau}A_b &= -\lambda_t^2 A_t - 6\lambda_b^2 A_b - \lambda_\tau^2 A_\tau + \frac{7}{15}g_1^2 M_1 + 3g_2^2 M_2 + \frac{16}{3}g_3^2 M_3 \\
\frac{d}{d\tau}A_\tau &= -3\lambda_b^2 A_b - 4\lambda_\tau^2 A_\tau + \frac{9}{5}g_1^2 M_1 + 3g_2^2 M_2 \\
\frac{d}{d\tau}B &= -3\lambda_t^2 A_t - 3\lambda_b^2 A_b - \lambda_\tau^2 A_\tau + \frac{3}{5}g_1^2 M_1 + 3g_2^2 M_2
\end{aligned} \tag{54}$$

where

$$\begin{aligned}
X_t^A &\equiv m_{\widetilde{Q}}^2 + m_{\tilde{t}}^2 + \mu_U^2 + A_t^2 \equiv X_t + A_t^2 \\
X_b^A &\equiv m_{\widetilde{Q}}^2 + m_{\tilde{b}}^2 + \mu_D^2 + A_b^2 \equiv X_b + A_b^2 \\
X_\tau^A &\equiv m_{\widetilde{L}}^2 + m_{\tilde{\tau}}^2 + \mu_D^2 + A_\tau^2 \equiv X_\tau + A_\tau^2
\end{aligned} \tag{55}$$

and $S = S_1 + S_2 + S_3$ where $S_3 = -\mu_Z^2 - \mu_A^2 - 2m_{\tilde{t}}^2 + m_{\tilde{b}}^2 + m_{\widetilde{Q}}^2 + m_{\tilde{\tau}}^2 - m_{\widetilde{L}}^2$ and $S_{1,2} = -2m_{\tilde{u},\tilde{c}}^2 + m_{\tilde{d},\tilde{s}}^2 + m_{\widetilde{Q}_{1,2}}^2 + m_{\tilde{e},\tilde{\mu}}^2 - m_{\widetilde{L}_{1,2}}^2$. S evolves according to $\frac{d}{d\tau}S = b_1 g_1^2 S$, and therefore satisfies the useful relation

$$-b_1 \int_{\tau_G}^{\tau} g_1^2 S = \left[1 - \frac{g_1^2(\tau)}{g_G^2}\right] S_G \,. \tag{56}$$

The gaugino masses are given by $M_i = M_{1/2}(g_i^2/g_G^2)$ where $M_{1/2}$ and g_G are the unified GUT-scale gaugino mass gauge coupling, respectively. The μ parameter in the superpotential runs according to

$$\frac{d}{d\tau}\mu = \left(-3\lambda_t^2 - 3\lambda_b^2 - \lambda_\tau^2 + \frac{3}{5}g_1^2 + 3g_2^2\right)\frac{1}{2}\mu\,. \tag{57}$$

Finally, the evolution of the gauge couplings is given by $\frac{d}{d\tau}g_i^2 = b_i g_i^4$ where $b_1 = -33/5$, $b_2 = -1$ and $b_3 = 3$ [note that we always use g_1 normalized as an SU(5) coupling], while the Yukawa couplings evolve according to

$$\begin{aligned}
\frac{d}{d\tau}\lambda_t^2 &= \left(-6\lambda_t^2 - \lambda_b^2 + \frac{13}{15}g_1^2 + 3g_2^2 + \frac{16}{3}g_3^2\right)\lambda_t^2 \\
\frac{d}{d\tau}\lambda_b^2 &= \left(-\lambda_t^2 - 6\lambda_b^2 - \lambda_\tau^2 + \frac{7}{15}g_1^2 + 3g_2^2 + \frac{16}{3}g_3^2\right)\lambda_b^2 \\
\frac{d}{d\tau}\lambda_\tau^2 &= \left(-3\lambda_b^2 - 4\lambda_\tau^2 + \frac{9}{5}g_1^2 + 3g_2^2\right)\lambda_\tau^2\,.
\end{aligned} \tag{58}$$

We now present the solution to the RG equations for the dimensionful parameters in terms of integrals over the dimensionless ones, namely the gauge and Yukawa couplings. Notice that μ renormalizes multiplicatively, in fact by a factor of order unity, and that is does not enter into the RG equations of the other mass parameters. For this reason, we may just as well treat μ at the electroweak scale as the fundamental parameter, and thus we will not need to refer to its GUT-scale value or its RG evolution.

The RG equations for the A parameters take the form

$$\frac{d}{d\tau}\vec{A} = H\vec{A} + M_{1/2}\vec{G}_A\,, \quad H \equiv -\begin{pmatrix} 6\lambda_t^2 & \lambda_b^2 & 0 \\ \lambda_t^2 & 6\lambda_b^2 & \lambda_\tau^2 \\ 0 & 3\lambda_b^2 & 4\lambda_\tau^2 \end{pmatrix} \tag{59}$$

where $\vec{G}_A = \left(\frac{13}{15}g_1^4 + 3g_2^4 + \frac{16}{3}g_3^4, \frac{7}{15}g_1^4 + 3g_2^4 + \frac{16}{3}g_3^4, \frac{9}{5}g_1^4 + 3g_2^4\right)/g_G^2$. The solution is given in terms of the "time"-ordered exponential of the integral of the matrix H,

$$\mathcal{H} \equiv T\left(\exp\int^\tau H\, d\tau'\right) \tag{60}$$

which satisfies $\frac{d}{d\tau}\mathcal{H} = H\,\mathcal{H}$. It may easily be computed numerically or estimated analytically. For example, with a GUT scale of $2.5\times10^{16}\,\mathrm{GeV}$ and $\alpha_G \simeq 1/24$, it is approximately given by

$$\begin{array}{ccc}
\lambda_b^G = \lambda_t^G = 0.6: & \lambda_b^G = \lambda_t^G = 1.0: & 2\lambda_b^G = \lambda_t^G = 1.0: \\
\mathcal{H} = & \mathcal{H} = & \mathcal{H} = \\
\begin{pmatrix} .239 & -.056 & .005 \\ -.061 & .278 & -.048 \\ .029 & -.290 & .610 \end{pmatrix} & \begin{pmatrix} .115 & -.042 & .007 \\ -.048 & .160 & -.050 \\ .039 & -.273 & .430 \end{pmatrix} & \begin{pmatrix} .103 & -.041 & .003 \\ -.073 & .372 & -.048 \\ .034 & -.267 & .670 \end{pmatrix}.
\end{array} \tag{61}$$

The trilinear couplings are then related to their GUT-scale values and to the gaugino mass at the GUT scale via

$$\vec{A} = \mathcal{H}\vec{A}_G + M_{1/2}\mathcal{H}\int_{\tau_G}^{\tau} \mathcal{H}^{-1}\vec{G}_A d\tau' \,. \tag{62}$$

The coefficient of $\vec{A}_G$ is typically an order of magnitude smaller than that of $M_{1/2}$ in the solutions for A_t and A_b, so we will often assume that $A_{t,b}$ are essentially determined at the electroweak scale by the gaugino masses. In the maximally symmetric case, both $M_{1/2}$ and $\vec{A}_G$ are negligible, while without the R symmetry the $M_{1/2}$ contribution is large and small effects due to $\vec{A}_G$ do not alter any conclusion substantially. (Of course one could imagine a scenario with $\vec{A}_G$ much larger than the gaugino mass and tuned to a particular value, for example to force $A_t \to 0$ and thus suppress the rate for $b \to s\gamma$, but we will not pursue this further. Such a tuning is implicitly included in $\epsilon_{b\to s\gamma}$.) We include approximate numerical expressions for $\vec{A}$ at the end of this appendix.

The RG equations (53) for the soft-breaking masses are now easily solved by noting that on their right-hand sides there are only three homogeneous driving terms, the X_i, for seven mass parameters. By taking linear combinations of the seven masses, specifically

$$\begin{pmatrix} X_t \\ X_b \\ X_\tau \\ I_1 \\ I_2 \\ I_3 \\ I_4 \end{pmatrix} = \begin{pmatrix} -1/2 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1/2 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1/2 & 1 & 0 & 0 & 0 & 1 & 1 \\ -1 & 0 & -19/4 & -7/4 & 7/2 & 0 & 0 \\ 0 & 0 & -1/2 & -1/2 & 1 & 0 & 0 \\ -1 & -1 & -4 & -1 & 5 & 0 & 1 \\ 1/2 & 1/2 & 3/2 & 0 & -3/2 & -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} \mu_Z^2 \\ \mu_A^2 \\ m_{\tilde{t}}^2 \\ m_{\tilde{b}}^2 \\ m_{\widetilde{Q}}^2 \\ m_{\tilde{\tau}}^2 \\ m_{\widetilde{L}}^2 \end{pmatrix} \tag{63}$$

or

$$\begin{pmatrix} \vec{X} \\ \vec{I} \end{pmatrix} \equiv M\,\vec{\mu} \tag{64}$$

we may write the RG equation of $\vec{\mu}$ in the form

$$\begin{aligned} \frac{d}{d\tau}\begin{pmatrix} \vec{X} \\ \vec{I} \end{pmatrix} &= \begin{pmatrix} H & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \vec{X} \\ \vec{I} \end{pmatrix} - \begin{pmatrix} 0 \\ \vec{v}_S \end{pmatrix} b_1 g_1^2 S \\ &\quad + \begin{pmatrix} H & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \vec{A}^2 \\ 0 \end{pmatrix} + \begin{pmatrix} \vec{G} \\ \vec{F} \end{pmatrix} M_{1/2}^2 \end{aligned} \tag{65}$$

where $\vec{v}_S \equiv (-\frac{25}{66}, -\frac{1}{33}, -\frac{1}{3}, \frac{5}{22})$ and $\vec{G}$ and $\vec{F}$ may be expressed in terms of g_i^6/g_G^4 using Eqs. (53). The solution to these equations is straightforward [recalling also Eq. (56)]:

$$\begin{aligned} \vec{\mu} \;=\; & M^{-1}\left[\begin{pmatrix} \mathcal{H}\vec{X}_G \\ \vec{I}_G \end{pmatrix} + \begin{pmatrix} 0 \\ \vec{v}_S \end{pmatrix}\left(1 - \frac{g_1^2}{g_G^2}\right) S_G + \right. \\ & \left. M_{1/2}^2 \begin{pmatrix} \mathcal{H}\int^\tau \mathcal{H}^{-1}\left\{\vec{G} + H\left[\mathcal{H}\left(\frac{\vec{A}_G}{M_{1/2}} + \int^{\tau'} \mathcal{H}^{-1}\vec{G}_A\, d\tau''\right)\right]^2\right\} d\tau' \\ \int^\tau \vec{F}\, d\tau' \end{pmatrix}\right]. \end{aligned} \tag{66}$$

Assuming once again unification at the above values of the GUT scale and α_G, and ignoring the A_G contribution, yields for the coefficient vector of $M_{1/2}^2 \simeq 1.6 m_{\widetilde{W}}^2$:

$$\begin{array}{ccc} \lambda_b^G = \lambda_t^G = 0.6: & \lambda_b^G = \lambda_t^G = 1.0: & 2\lambda_b^G = \lambda_t^G = 1.0: \\ M^{-1}\begin{pmatrix} 7.32 \\ 7.64 \\ -1.83 \\ -18.17 \\ 0.46 \\ 2.73 \\ -0.46 \end{pmatrix} = \begin{pmatrix} 5.45 \\ -5.23 \\ 4.77 \\ 4.87 \\ 5.28 \\ 0.15 \\ 0.53 \end{pmatrix} & M^{-1}\begin{pmatrix} 6.84 \\ 7.22 \\ -1.96 \\ -18.17 \\ 0.46 \\ 2.73 \\ -0.46 \end{pmatrix} = \begin{pmatrix} 5.87 \\ -5.61 \\ 4.63 \\ 4.75 \\ 5.15 \\ 0.17 \\ 0.54 \end{pmatrix} & M^{-1}\begin{pmatrix} 6.71 \\ 8.23 \\ -1.51 \\ -18.17 \\ 0.46 \\ 2.73 \\ -0.46 \end{pmatrix} = \begin{pmatrix} 6.19 \\ -5.25 \\ 4.53 \\ 5.11 \\ 5.27 \\ 0.13 \\ 0.52 \end{pmatrix} . \end{array} \tag{67}$$

Finally, the RG equation for B,

$$\frac{d}{d\tau} B = \vec{H}_B \cdot \vec{A} + G_B M_{1/2} \tag{68}$$

where $\vec{H}_B = (-3\lambda_t^2, -3\lambda_b^2, -\lambda_\tau^2)$ and $G_B = (\frac{3}{5} g_1^4 + 3 g_2^4)/g_G^2$, is solved by simply integrating over the gaugino and A contributions:

$$B = B_G + \left(\int^\tau \vec{H}_B \mathcal{H} \, d\tau' \right) \vec{A}_G + M_{1/2} \int^\tau \left(\vec{H}_B \mathcal{H} \int^{\tau'} \mathcal{H}^{-1} \vec{G}_A \, d\tau'' + G_B \right) d\tau' \, . \tag{69}$$

Under the same unification assumptions as before, we obtain for $\lambda_b^G = \lambda_t^G = 0.6$: $B = B_G - (0.36, 0.33, 0.08)\vec{A}_G - 1.03\, M_{1/2}$; for $\lambda_b^G = \lambda_t^G = 1.0$: $B = B_G - (0.41, 0.36, 0.11)\vec{A}_G - 1.25\, M_{1/2}$; and for $2\lambda_b^G = \lambda_t^G = 1.0$: $B = B_G - (0.43, 0.28, 0.07)\vec{A}_G - 1.08\, M_{1/2}$.

The various integrals involving the gauge and Yukawa couplings may be approximately evaluated analytically, since the evolution of the g_i is known and simple while the λ_i may be approximated in various ways, in particular near the fixed-point regime. However, for our purposes the semi-analytic forms presented above are sufficient. To get a feel for the results, we can evaluate the integrals numerically. Using the same unification scale and gauge couplings as above, inserting the initial conditions dictated by SU(5) symmetry, and setting $A_{t,b,\tau} = A_G$ at the GUT scale but neglecting (for ease of presentation) the small contributions of A_G to the soft-breaking masses, we obtain the following (approximate) explicit solutions:

$$\begin{array}{rcl} & & \lambda_b^G = \lambda_t^G = 0.6: \\ m_Z^2 &=& 5.45\, M_{1/2}^2 - 1.29\, M_{\mathbf{10}_H}^2 + 1.41\, M_{\mathbf{16}_3}^2 + 0.38\, M_X^2 - 0.11\, M_{\mathrm{SU}(5)}^2 - 2\mu^2 \\ m_A^2 &=& -5.23\, M_{1/2}^2 + 1.25\, M_{\mathbf{10}_H}^2 - 1.50\, M_{\mathbf{16}_3}^2 + 0.05\, M_{\mathrm{SU}(5)}^2 + 2\mu^2 \\ m_{\tilde{t}}^2 &=& 4.77\, M_{1/2}^2 - 0.24\, M_{\mathbf{10}_H}^2 + 0.53\, M_{\mathbf{16}_3}^2 + 0.09\, M_X^2 + 0.01\, M_{\mathrm{SU}(5)}^2 \\ m_{\tilde{b}}^2 &=& 4.87\, M_{1/2}^2 - 0.21\, M_{\mathbf{10}_H}^2 + 0.58\, M_{\mathbf{16}_3}^2 - 0.29\, M_X^2 - 0.51\, M_{\mathrm{SU}(5)}^2 \\ m_{\widetilde{Q}}^2 &=& 5.28\, M_{1/2}^2 - 0.22\, M_{\mathbf{10}_H}^2 + 0.56\, M_{\mathbf{16}_3}^2 + 0.10\, M_X^2 + 0.23\, M_{\mathrm{SU}(5)}^2 \\ m_{\tilde{\tau}}^2 &=& 0.15\, M_{1/2}^2 - 0.17\, M_{\mathbf{10}_H}^2 + 0.66\, M_{\mathbf{16}_3}^2 + 0.12\, M_X^2 + 0.36\, M_{\mathrm{SU}(5)}^2 \end{array} \tag{70}$$

$$
\begin{aligned}
m^2_{\widetilde{L}} &= 0.53\, M^2_{1/2} - 0.08\, M^2_{\mathbf{10}_H} + 0.83\, M^2_{\mathbf{16}_3} - 0.31\, M^2_X - 0.64\, M^2_{\mathrm{SU}(5)} \\
A_t &= 0.19\, A_G + 2.2 \quad M_{1/2} \\
A_b &= 0.17\, A_G + 2.3 \quad M_{1/2} \\
A_\tau &= 0.35\, A_G - 0.13\, M_{1/2} \\
B &= B_G - 0.76\, A_G - 1.0\, M_{1/2} \\
\mu &= 0.65\, \mu^G\,;
\end{aligned}
$$

$$
\lambda^G_b = \lambda^G_t = 1.0:
$$

$$
\begin{aligned}
m^2_Z &= 5.87\, M^2_{1/2} - 1.20\, M^2_{\mathbf{10}_H} + 1.61\, M^2_{\mathbf{16}_3} + 0.38\, M^2_X + 0.05\, M^2_{\mathrm{SU}(5)} - 2\mu^2 \\
m^2_A &= -5.61\, M^2_{1/2} + 1.14\, M^2_{\mathbf{10}_H} - 1.72\, M^2_{\mathbf{16}_3} + 0.06\, M^2_{\mathrm{SU}(5)} + 2\mu^2 \\
m^2_{\tilde{t}} &= 4.63\, M^2_{1/2} - 0.27\, M^2_{\mathbf{10}_H} + 0.45\, M^2_{\mathbf{16}_3} + 0.09\, M^2_X - 0.05\, M^2_{\mathrm{SU}(5)} \\
m^2_{\tilde{b}} &= 4.75\, M^2_{1/2} - 0.23\, M^2_{\mathbf{10}_H} + 0.54\, M^2_{\mathbf{16}_3} - 0.29\, M^2_X - 0.47\, M^2_{\mathrm{SU}(5)} \\
m^2_{\widetilde{Q}} &= 5.14\, M^2_{1/2} - 0.25\, M^2_{\mathbf{10}_H} + 0.50\, M^2_{\mathbf{16}_3} + 0.10\, M^2_X + 0.23\, M^2_{\mathrm{SU}(5)} \\
m^2_{\tilde{\tau}} &= 0.17\, M^2_{1/2} - 0.23\, M^2_{\mathbf{10}_H} + 0.54\, M^2_{\mathbf{16}_3} + 0.12\, M^2_X + 0.41\, M^2_{\mathrm{SU}(5)} \qquad (71) \\
m^2_{\widetilde{L}} &= 0.54\, M^2_{1/2} - 0.12\, M^2_{\mathbf{10}_H} + 0.77\, M^2_{\mathbf{16}_3} - 0.31\, M^2_X - 0.62\, M^2_{\mathrm{SU}(5)} \\
A_t &= 0.08\, A_G + 2.0 \quad M_{1/2} \\
A_b &= 0.06\, A_G + 2.0 \quad M_{1/2} \\
A_\tau &= 0.20\, A_G - 0.22\, M_{1/2} \\
B &= B_G - 0.88\, A_G - 1.2\, M_{1/2} \\
\mu &= 0.43\, \mu^G\,;
\end{aligned}
$$

$$
2\lambda^G_b = \lambda^G_t = 1.0:
$$

$$
\begin{aligned}
m^2_Z &= 6.19\, M^2_{1/2} - 1.15\, M^2_{\mathbf{10}_H} + 1.69\, M^2_{\mathbf{16}_3} + 0.38\, M^2_X + 0.03\, M^2_{\mathrm{SU}(5)} - 2\mu^2 \\
m^2_A &= -5.25\, M^2_{1/2} + 1.24\, M^2_{\mathbf{10}_H} - 1.53\, M^2_{\mathbf{16}_3} - 0.07\, M^2_{\mathrm{SU}(5)} + 2\mu^2 \\
m^2_{\tilde{t}} &= 4.53\, M^2_{1/2} - 0.28\, M^2_{\mathbf{10}_H} + 0.44\, M^2_{\mathbf{16}_3} + 0.09\, M^2_X - 0.04\, M^2_{\mathrm{SU}(5)} \\
m^2_{\tilde{b}} &= 5.11\, M^2_{1/2} - 0.18\, M^2_{\mathbf{10}_H} + 0.64\, M^2_{\mathbf{16}_3} - 0.29\, M^2_X - 0.54\, M^2_{\mathrm{SU}(5)} \\
m^2_{\widetilde{Q}} &= 5.27\, M^2_{1/2} - 0.23\, M^2_{\mathbf{10}_H} + 0.54\, M^2_{\mathbf{16}_3} + 0.10\, M^2_X + 0.19\, M^2_{\mathrm{SU}(5)} \\
m^2_{\tilde{\tau}} &= 0.13\, M^2_{1/2} - 0.15\, M^2_{\mathbf{10}_H} + 0.70\, M^2_{\mathbf{16}_3} + 0.12\, M^2_X + 0.34\, M^2_{\mathrm{SU}(5)} \qquad (72) \\
m^2_{\widetilde{L}} &= 0.52\, M^2_{1/2} - 0.07\, M^2_{\mathbf{10}_H} + 0.85\, M^2_{\mathbf{16}_3} - 0.31\, M^2_X - 0.65\, M^2_{\mathrm{SU}(5)} \\
A_t &= 0.07\, A_G + 1.9 \quad M_{1/2} \\
A_b &= 0.25\, A_G + 2.4 \quad M_{1/2} \\
A_\tau &= 0.44\, A_G - 0.01\, M_{1/2} \\
B &= B_G - 0.78\, A_G - 1.1\, M_{1/2} \\
\mu &= 0.57\, \mu^G\,.
\end{aligned}
$$

Appendix B: The flat direction ϕ_2

In this appendix we discuss in more detail the constraints implied by the flat direction ϕ_2 of Sec. 6.1. As discussed in that section, at high enough scales ($\Lambda > \Lambda_{\rm HIGH}$) the direction may be stabilized by nonrenormalizable operators regardless of the sign of m_2^2. At a lower scale, such operators are ineffective, and a negative m_2^2 leads to an instability unless the linear term $(m_{\widetilde{L}}^2 + m_{\widetilde{Q}}^2 + m_{\bar{b}}^2)\left|\frac{\mu}{\lambda_b}\phi_2\right| \equiv m_3^2\left|\frac{\mu}{\lambda_b}\phi_2\right|$ in the scalar potential is significant at such a scale. We need to estimate the scale $\Lambda_{\rm LOW}$ down to which this linear term may be ignored, and therefore above which $m_2^2(\Lambda) > 0$ must be enforced.

If $m_2^2(\Lambda) > 0$ for all Λ between $M_{\rm GUT}$ and m_S, there is no instability. If $m_2^2(\Lambda) < 0$ for some $\Lambda > \Lambda_{\rm HIGH}$, the dangerous minimum in the potential is only at field values $\phi_2 \ll \Lambda_{\rm HIGH}$ (by construction of $\Lambda_{\rm HIGH}$), so to see if it is a true minimum we must run to lower scales. If $m_2^2(\Lambda) < 0$ when we reach $\Lambda = \Lambda_{\rm HIGH}$, then the true minimum is at $\phi_2 \sim \Lambda_{\rm HIGH}$ and leads to unacceptable symmetry breaking. If $m_2^2(\Lambda_{\rm HIGH}) > 0$, there is no dangerous minimum at that scale, and we should continue running to lower scales. If $m_2^2(\Lambda)$ gets to zero at a scale Λ_C above the scale $\Lambda_{\rm LOW}$ (to be determined below) so the linear term in the potential may be ignored, we must minimize the full 1-loop effective potential along the flat direction [35]. At 1-loop order we parametrize the flat direction by a field ϕ_2 with zero anomalous dimension, so $\langle H_u\rangle = \phi_2 z_u^{-1/2}$ and $\langle L\rangle = \phi_2 z_L^{-1/2}$, where $z_{u,L}$ are wavefunction renormalization coefficient functions satisfying the RG equations $\partial \ln z_{u,L}/\partial \ln(\phi_2/\Lambda) \equiv \gamma_{u,L}$. (Notice that $\langle H_U\rangle \neq \langle L\rangle$ because the D-flatness condition which determines the VEVs of H_u and L is corrected by 1-loop wavefunction renormalizations.) Then in leading-$\ln(\phi_2/\Lambda)$ approximation the full 1-loop effective potential (neglecting the linear term) is completely determined by the RG equation for m_2^2 calculated using Appendix A:

$$\begin{aligned} V_0 + V_1 &= m_2^2|\phi_2|^2 + \frac{1}{8\pi^2}\left(3\lambda_t^2 X_t + \lambda_\tau^2 X_\tau - \frac{6}{5}g_1^2 M_1^2 - 6 g_2^2 M_2^2\right)|\phi_2|^2 \ln\left|\frac{\phi_2}{\Lambda}\right| \\ &\equiv m_2^2|\phi_2|^2 + \Delta m_2^2 |\phi_2|^2 \ln\left|\frac{\phi_2}{\Lambda}\right| . \end{aligned} \tag{73}$$

At the scale Λ_C, where $m_2^2 = 0$, the above potential has the well-known Coleman-Weinberg minimum at $\langle\phi_2\rangle \sim \Lambda_C$. Moreover the vacuum energy at that minimum is $\sim -\Delta m_2^2 \Lambda_C^2$, which is parametrically much below the usual electroweak vacuum energy $-\mathcal{O}(m_Z^4/g^2)$. If Λ_C and therefore $\langle\phi_2\rangle$ get too low, the linear term $m_3^2\left|\frac{\mu\phi_2}{\lambda_b}\right|$ in the potential dominates, and the unwanted minimum disappears. To get a rough estimate of when this happens, we just add the linear term to Eq. (73) and again minimize at $\Lambda = \Lambda_C$. We find that the dangerous minimum is eliminated because of the linear term when $\Lambda_C \lesssim \frac{1}{2}e^{3/2}(m_3^2/\Delta m_2^2)\mu/\lambda_b$. Thus, $m_2^2(\Lambda)$ must be prevented from vanishing only above scales of order

$$\Lambda_{\rm LOW} \sim \frac{m_3^2\mu}{\Delta m_2^2 \lambda_b} \sim \frac{2\pi}{\alpha}\frac{\mu}{\lambda_b}, \tag{74}$$

where the quantities on the right hand side are evaluated at Λ_C and α is a combination of gauge and Yukawa coupling strengths. The above is just an estimate. For instance, 2-

loop RG effects and finite parts in the 1-loop potential modify the numerical prefactor on the right-hand side by $\mathcal{O}(1)$. Moreover, notice that along ϕ_2 there is a hierarchy between the scale of SU(2) breaking ($\sim \langle H_u \rangle$) and that of color breaking ($\sim \langle Q \rangle$). Therefore we expect the 1-loop corrections to m_3^2, which we haven't included in our estimate, to be of order $\alpha \ln(\langle Q \rangle / \Lambda_C) \sim \alpha \ln(1/\alpha)$.

Appendix C: Approximations in the effective potential

Throughout the paper our analysis has been based on the 1-loop-RG improved tree level scalar potential $V_0(\Lambda)$. In this appendix we discuss the possible relevance of a more accurate treatment which would include the full 1-loop effective potential, 1-loop threshold corrections and 2-loop running of the soft SUSY-breaking parameters. We will show that all these refinements cannot change the basic conclusions of our study.

Let us first discuss the use of the 1-loop effective potential. This becomes necessary when the tree-level potential is almost flat (or even unbounded) along some direction in field space. Then the quantum correction stabilizes the potential at large field strength. Minimization of the potential then yields a vacuum expectation value for that field of order the renormalization scale Λ at which the potential becomes flat (dimensional transmutation)[35]. (Of course we are assuming that at very high energy scales the tree potential is bounded from below). In our study, as discussed in Sec. 6.1, we need only worry about the two flat directions ϕ_1 and ϕ_2. Hence the full 1-loop effective potential is only relevant for those parameter ranges when m_1^2 or m_2^2 become very small, namely very near the $m_1^2 = 0$ or $m_2^2 = 0$ planes in the space of GUT parameters. Thus a more correct and involved calculation would only change the margins of the allowed parameter space but could not significantly alter any conclusions.

The remaining improvements are given by (*i*) GUT-scale thresholds, (*ii*) SUSY-scale thresholds, and (*iii*) 2-loop running. The first type of effects are model-dependent and have essentially been encompassed by our discussions of the various boundary conditions. Let us then turn to the effects of the superpartner thresholds on the allowed (usually triangular) regions of SO(10) parameter space—the extension to more general initial parameters will be obvious. The first question is where to stop the running. In the plots in Fig. 7 the running has been stopped at $\Lambda = m_Z$, even though in the typical scenarios the superpartners decouple closer to 1 TeV. A more exact analysis would then add 1-loop threshold corrections to the various mass parameters. (In the notation of Ref. [37], such threshold effects would appear as terms in the 1-loop effective potential.) These corrections are roughly proportional to $\alpha m_S \ln(m_S/m_Z)$, where m_S collectively denotes the low energy values of the MSSM mass parameters and α is the appropriate gauge or Yukawa coupling strength. In the absence of strong premature focusing (for example, λ_G well below 1.2 in the SO(10) scenario), when all the superpartner masses are comparable, the effect on the triangle plots is just to move the triangle boundaries by an amount which is roughly $\mathcal{O}[\alpha \ln(m_S/m_Z)]$ times the size of the triangle itself. The size of the allowed region in the GUT parameter space is only slightly changed, and the same arguments we have made can be applied to the slightly shifted GUT parameter ranges. If there is strong premature focusing—that is, if λ_G is just below its maximal value, the al-

lowed region of parameter space is very small, and some particles have masses $m^2_{0,L}$ well below the rest of the superpartner masses $m^2_{0,H}$—then the threshold corrections could be significant when $\alpha m^2_{0,H} \ln(m^2_{0,H}/m_Z) \sim m^2_{0,L}$. But their only relevant effect would be to slightly shift the maximal value of λ_G, so once again no conclusions are qualitatively altered. Notice that it was crucial to establish that the corrections are proportional to the *electroweak-scale* values of the masses, which can be much smaller that the GUT values when any focusing is relevant. The effects of 2-loop running, however, are in general proportional to the value of the masses at higher scales: we expect corrections $\sim \alpha M_0^2$ to the low energy values of the masses (recall that M_0 is a typical soft mass at $M_{\rm GUT}$). When λ_G is large and $\lambda_t^G = \lambda_b^G$, the low-energy masses in the allowed regions of parameter space are focused to small values m_0^2. When the focusing is strong enough that $m_0^2 \sim \alpha M_0^2$, some effects of 2-loop running are large. For example, a value of λ_G which leads to a small but finite allowed triangle with 1-loop running could lead to vanishing allowed area using 2-loop running. So when there is large focusing the more exact upper bound on λ_G could shift somewhat—but because of the fixed-point evolution of λ_t at large λ_G, the corresponding values of m_t will not change much. We therefore expect that all our conclusions are robust.

References

[1] H. Georgi and S.L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

[2] H. Georgi, in *AIP Conference Proceedings* **23**, ed. C. Carlson (AIP Press, 1975).

[3] J.C. Pati and A. Salam, Phys. Rev. D **10**, 275 (1974).

[4] R.N. Mohapatra and J.C. Pati, Phys. Rev. D **11**, 2558 (1975); G. Senjanović and R.N. Mohapatra, Phys. Rev. D **12**, 1502 (1975).

[5] M. Veltman, Acta Phys. Pol. **B12**, 437 (1981); L. Maiani, Proc. Summer School of Gif-sur-Ivette (Paris 1980) p.3; E. Witten, Nucl. Phys. **B188**, 513 (1981).

[6] H. Georgi, H. Quinn, and S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974); S. Dimopoulos, S. Raby, and F. Wilczek, Phys. Rev. D **24**, 1681 (1981); S. Dimopoulos and H. Georgi, Nucl. Phys. **B193**, 150 (1981); L. Ibañez and G.G. Ross, Phys. Lett. B **105**, 439 (1981); M.B. Einhorn and D.R.T. Jones, Nucl. Phys. **B196**, 475 (1982); W.J. Marciano and G. Senjanovic, Phys. Rev. D **25**, 3092 (1982).

[7] U. Amaldi et al., Phys. Rev. D **36**, 1385 (1987); G. Costa et al., Nucl. Phys. **B297**, 244 (1988).

[8] P. Langacker and M.-X. Luo, Phys. Rev. **D44**, 817 (1991); U. Amaldi, W. de Boer, and H. Fürstenau, Phys. Lett. **B260**, 447 (1991); J. Ellis, S. Kelley, and D.V. Nanopoulos, Phys. Lett. **B260**, 131 (1991); F. Anselmo, L. Cifarelli, A. Peterman, and A. Zichichi,

Nuovo Cimento **104A**, 1817 (1991); W.J. Marciano, Ann. Rev. Nucl. Part Phys. **41**, 469 (1992).

[9] M. Chanowitz, J. Ellis, and M.K. Gaillard, Nucl. Phys. **B135**, 66 (1978).

[10] L.E. Ibañez and C. Lopez, Phys. Lett. B **126**, 54 (1983) and Nucl. Phys. **B233**, 511 (1984); H. Arason *et al.* [13]; L.J. Hall and U. Sarid, Phys. Lett. B **271**, 138 (1991); S. Kelley *et al.* [13].

[11] L.E. Ibañez and G.G. Ross, Phys. Lett. B **110**, 215 (1982); K. Inoue, A. Kakuto, H. Komatsu and S. Takeshita, Prog. Theor. Phys. **68**, 927 (1982) and *ibid.* **71** 413 (1984); J. Ellis, D.V. Nanopoulos and K. Tamvakis, Phys. Lett. B **121** 123 (1983); L.E. Ibañez, Nucl. Phys. **B218**, 514 (1983).

[12] L. Alvarez-Gaumè, J. Polchinski and M.B. Wise, Nucl. Phys. **B221**, 495 (1983).

[13] G.F. Giudice and G. Ridolfi, Z. Phys. C **41**, 447 (1988); M. Olechowski and S. Pokorski, Phys. Lett. B **214**, 393 (1988); P.H. Chankowski, Phys. Rev. D **41**, 2877 (1990); M. Drees and M.M. Nojiri, Nucl. Phys. **B369**, 54 (1992); B. Ananthanarayan, G. Lazarides, and Q. Shafi, Phys. Lett. B **300**, 245 (1993); H. Arason, D.J. Castaño, B.E. Keszthelyi, S. Mikaelian, E.J. Piard, P. Ramond, and B.D. Wright, Phys. Rev. Lett. **67**, 2933 (1991); S. Kelley, J.L. Lopez, and D.V. Nanopoulos, Phys. Lett. B **274**, 387 (1992); V. Barger, M.S. Berger, and P. Ohmann, Phys. Rev. D **47**, 1093 (1993); M. Carena, S. Pokorski and C.E.M. Wagner, Nucl. Phys. **B406**, 59 (1993).

[14] M. Carena, M. Olechowski, S. Pokorski and C.E.M. Wagner, Nucl. Phys. **B426**, 269 (1994).

[15] L.J. Hall, R. Rattazzi and U. Sarid, Phys. Rev. D **50**, 7048 (1994).

[16] R. Rattazzi, U. Sarid and L.J. Hall, SU-ITP-94-15, RU-94-37, *Proceedings of the Second IFT Workshop on Yukawa Couplings and the Origins of Mass* (1994).

[17] B. Ananthanarayan, G. Lazarides, and Q. Shafi, Phys. Rev. D **44**, 1613 (1991).

[18] S. Chaudhury, S.W. Chung, G. Hockney and J. Lykken, PRINT-94-0170 (FERMILAB), Aug 1994, hep-th/9409151. Talk given at 1994 Meeting of the American Physical Society, Division of Particles and Fields (DPF 94), Albuquerque. Also J. Lykken, private communication.

[19] F. Vissani and A. Yu. Smirnov, Phys. Lett. B **341**, 173 (1994).

[20] A. Brignole, H. Murayama and R. Rattazzi, Phys. Lett. B **335**, 345 (1994).

[21] A.E. Nelson and L. Randall, Phys. Lett. B **316**, 516 (1993).

[22] R. Hempfling, Phys. Rev. D **49**, 6168 (1994).

[23] E.C. Poggio, H.R. Quinn, and S. Weinberg, Phys. Rev. D **13**, 1958 (1976); M.A. Shifman, A.I. Vainshtein, and V.I. Zakharov, Nucl. Phys. **B147**, 385 and 448 (1979); B. Guberina, R. Meckbach, R.D. Peccei, and R. Rückl, Nucl. Phys. **B184**, 476 (1981).

[24] J. Gasser and H. Leutwyler, Phys. Rep. **87**, 77 (1982).

[25] N. Polonsky, Preprint UPR-0595/T (May 1994), published in SUSY-94 Workshop, p. 438.; M. Carena, S. Dimopoulos, C.E.M. Wagner and S. Raby, Preprint CERN-TH/95/53.

[26] DØ Collaboration (S. Abachi, et al.). FERMILAB-PUB-95-028-E, Mar 1995. 12pp. Submitted to Phys. Rev. Lett.; CDF Collaboration (F. Abe, et al.). FERMILAB-PUB-95-022-E, Mar 1995. 18pp.

[27] R. Garisto and J.N. Ng, Phys. Lett. **B315**, 372 (1993); M.A. Díaz, Phys. Lett. **B322**, 207 (1994); F.M. Borzumati, Z. Phys. C **63**, 291 (1994).

[28] M.S. Alam *et al.*, CLEO Collaboration, Phys. Rev. Lett. **74**, 2885 (1995).

[29] R. Rattazzi and U. Sarid, in preparation.

[30] S. Bertolini, F. Borzumati, A. Masiero and G. Ridolfi, Nucl. Phys. **B353**, 591 (1991); M. Misiak, Phys. Lett. B **269**, 161 (1991); R. Barbieri and G.F. Giudice, Phys. Lett. B **309**, 86 (1993).

[31] F. Zwirner, in *Physics and experiments with linear colliders*, edited by R. Orava, P. Eerola and M. Nordberg (World Scientific, Singapore, 1992), Vol. I, p. 309 [CERN Theoretical Report No. CERN-Th.6357/91], and references therein.

[32] M. Olechowski and S. Pokorski, Phys. Lett. B **344**, 201 (1995); F.M. Borzumati, M. Olechowski and S. Pokorski, Phys. Lett. B **349**, 311 (1995).

[33] A. Lleyda and C. Muñoz, Phys. Lett. B **317**, 82 (1993).

[34] G. Costa, F. Feruglio, F. Gabbiani and F. Zwirner, Nucl. Phys. **B286**, 325 (1987).

[35] S. Coleman and E. Weinberg, Phys. Rev. D **7**, 1888 (1973); S. Weinberg, Phys. Rev. D **7**, 2887 (1973).

[36] H. Komatsu, Phys. Lett. B **215**, 323 (1988).

[37] G. Gamberini, G. Ridolfi and F. Zwirner, Nucl. Phys. **B331**, 331 (1990).

[38] J.M. Frère, D.R.T. Jones and S. Raby, Nucl. Phys. **B222**, 11 (1983); J.-P. Derendinger and C.A. Savoy, Nucl. Phys. **B237**, 307 (1984).

[39] M. Drees, Phys. Lett. B **181**, 279 (1986); J.S. Hagelin and S. Kelley, Nucl. Phys. **B342**, 95 (1990); A.E. Faraggi, S. Kelley and D.V. Nanopoulos, Phys. Rev. D **45** 3272 (1992); Y. Kawamura, H. Murayama and M. Yamaguchi, Phys. Lett. B **324**, 52 (1994); R. Rattazzi, U. Sarid and L.J. Hall [16]; H.C. Cheng and L.J. Hall, Preprint LBL-35950; C. Kolda and S. Martin, Preprint UM-TH-95-08; R. Hempfling, Preprint DESY-94-078.

[40] Y. Kawamura, H. Murayama and M. Yamaguchi, Phys. Rev. D **51**, 1337 (1995).

[41] S. Dimopoulos and F. Wilczek, ITP Santa Barbara Preprint UM HE 81-71 (1981) and Proceedings Erice Summer School, Ed. A. Zichichi (1981); K.S. Babu and S.M. Barr, Phys. Rev. D **48**, 5354 (1993).

[42] S. Wolfram, Phys. Lett. **B82**, 65 (1979); C. B. Dover *et al.*, Phys. Rev. Lett. **42**, 1117 (1979); P. F. Smith *et al.*, Nucl. Phys. **B206**, 333 (1982).

[43] M. Drees and M. M. Nojiri, Phys. Rev. D **47**, 376 (1992).

[44] B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977).

[45] K. Griest and D. Seckel, Phys. Rev. D **43**, 3191

[46] S. Mizuta and M. Yamaguchi, Phys. Lett. B **298**, 120 (1993).

[47] V. Kaplunovsky and J. Louis, Phys. Lett. B **306**, 269 (1993).

[48] L.J. Hall, Lectures given at Winter School in Theoretical Physics, Mahabaleshwar, India, January 1984; Harvard University preprint HUTP-84/A012 (March 1984); Published in Indian Winter School 1984, p. 197.